Structural studies on enzymes involved in uronic acid polysaccharide degradation

A Thesis Submitted to the College of Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

ThirumalaiSelvi Ulaganathan, B.Tech.

2018

Permission to Use

In presenting this thesis, in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College under which this thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis.

Request for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Dr. Miroslaw Cygler
Department of Biochemistry
Health Sciences Building, 107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

ABSTRACT

Uronates are charged sugars that constitute a major component of glycosaminoglycans (heparin, heparin sulfate, hyaluronan etc.,) and cell wall polysaccharides of plants (pectate/ pectin) and marine algae (alginate, ulvan). Glycosaminoglycans (GAGs) participate in a variety of cellular functions including stabilization of the extracellular matrix, control of hydration and organism development, whereas pectate, alginate and ulvan represent a huge biomass with wide industrial applications. Bacteria that encounter GAGs and plant algal biomass express enzymes to degrade uronate polysaccharides as a source of nutrients. These enzymes are called polysaccharide lyases (PL) and degrade the uronate polysaccharide by a β-elimination mechanism. Polysaccharide lyase enzymes are known to be involved in microbial pathogenicity and gut microbiome interactions. In addition, PL enzymes serve as biotechnology tools for industrial biomass conversion.

The overall objective of this PhD project is to expand the existing structural and mechanistic knowledge of enzymes belonging to the PL families, with the emphasis on newly identified families for which no three-dimensional structures or catalytic mechanisms have been established. The first part of the thesis describes the structure and catalytic mechanism of **Heparinase III** from *Bacteroides thetaiotaomicron*. This enzyme specifically degrades the heparan sulfate GAG. The second part of the thesis concerns with the structure and catalytic mechanism of three newly identified lyases that degrade ulvan, the cell wall polysaccharide present in marine green algae. These ulvan lyases are; **LOR107** from *Alteromonas*, **PLSV3936** from *Pseudoalteromonas*, **NLR48** form *Nonlabens ulvanivorans*.

The methodologies used in this research included: (1) X-Ray crystallography to determine the atomic structure of the aforementioned enzymes and the enzyme/substrate complex, (2) site-directed mutagenesis of key residues involved in substrate binding or catalysis, (3) enzymatic characterization of designed mutants to obtain insight into the roles of the selected residues in the catalytic process.

These studies resulted in: (1) the structure of HepIII and the identification of conformational flexibility prevalent in this class of enzymes, (2) the first structures of three ulvan lyases, called LOR107, PLSV3936 and NLR48. The ulvan lyases LOR107 and PLSV3936 share the 7-bladed β-propeller fold and NLR48 has a β-jelly roll fold. Despite different structural scaffold, the β-elimination catalytic machinery is conserved among the families. However, the complex structure with the bound tetrasaccharide substrate reveals the difference in the active site and mode of substrate cleavage. Overall, the structures determined as part of this research provide structural templates for a large number of lyases classified into the four different PL families.

ACKNOWLEDGMENTS

I would like to thank my supervisor **Dr. Miroslaw Cygler** for providing me the opportunity to work in his group. I am grateful for his valuable mentorship and encouragement throughout my research.

I extend my sincere gratitude to my advisory committee members Dr. Jeremy Lee, Dr. Stanley Moore, Dr. David Sanders and Dr. Peter Howard. I thank Dr. Alisdair Boraston for acting as my external examiner.

I would like to thank the present and former members of Cygler's lab, Dr. Maia Cherney, Dr. Usha Nair, Dr. Michal Boniecki, Dr. Andrey Grishin, Dr. Ksenia Beyrakhova, Dr. Ivy Chung, Dr. Lei Li, Ben Flath, Caishuang Xu and Kevin Voth for the collegial atmosphere in the lab, productive scientific discussions and technical help.

I thank all our collaborators who facilitated my research to be fruitful. I thank the Protein Characterization facility (PCCF) for the access to various instruments and the CLS-CMCF beamline scientists.

I thank the Canadian National Committee for Crystallography and IUCr for awarding the Larry Calvert and young scientist travel award. I also thank the Department of Biochemistry, College of Medicine, College of graduate studies and research and other Science societies for the various travel awards.

I greatly appreciate the assistance of the members of Biochemistry department, throughout my graduate program.

My hearty thanks to my husband (Guru), sister (Ponni) and friends who provide the love, support and constant encouragement. Finally, I thank and dedicate this thesis to my parents and my teachers, whose hard work and faith in me has been instrumental in achieving all my success.

PREFACE

The findings reported in this thesis have been published in four research articles. All figures from the following publications have been used in this thesis.

Section 5.1

Ulaganathan, T., Shi, R., Yao, D., Gu, R.X., Garron, M.L., Cherney, M., Tieleman, D.P., Sterner, E., Li, G., Li, L. and Linhardt, R.J. (2017). Conformational flexibility of PL12 family heparinases: Structure and substrate specificity of heparinase III from Bacteroides thetaiotaomicron (BT4657). Glycob. 27(2), 176-187. doi.org/10.1093/glycob/cww096

Section 6.1

Ulaganathan, T., Helbert, W., Kopel, M., Banin, E., and Cygler, M. (2018). Structure–function analyses of a PL24 family ulvan lyase reveal key features and suggest its catalytic mechanism. J. Biol. Chem. 293(11), 4026-4036. doi: 10.1074/jbc.RA117.001642

Section 6.2

Ulaganathan, T., Boniecki, M. T., Foran, E., Buravenkov, V., Mizrachi, N., Banin, E., Helbert, and Cygler, M. (2017). New ulvan-degrading polysaccharide lyase family: structure and catalytic mechanism suggests convergent evolution of active site architecture. ACS. Chem. Biol. 12(5), 1269-1280. doi: 10.1021/acschembio.7b00126

Section 6.3

Ulaganathan, T., N., Banin, E., Helbert, and Cygler, M. (2018) Structural and functional characterization of PL28 family ulvan lyase NLR48 from Nonlabens ulvanivorans. J. Biol. Chem. jbc-RA118. doi: 10.1074/jbc.RA118.003659

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ChonABC	Chondroitinase ABC
ChonAC	Chondroitin AC lyase
ChonB	Chondroitinase B
CS	Chondroitin sulfate
CS-A	Chondroitin 4- sulfate/ Chondroitin sulfate A
CS-C	Chondroitin 6-sulfate or Chondroitin sulfate C
CsL	Chondroitin lyase
DEH	4-deoxy-L-erythro-5-hexoseulose uronic acid
DehR	DEH reductase
DS, CS-B	Dermatan sulfate
ECM	Extracellular Matrix
ED	Entner–Doudoroff pathway
GAGs	Glycosaminoglycans
GAL	Galactose
GalA	Galacturonic acid
GalNAc	N-acetylgalactosamine
GH	Glycoside hydrolase
GlcNAc	N-acetylglucosamine
Glu	Glucose
GlcA	Glucuronic acid
GST	Glutathione S-transferase
HA	Hyaluronan
HepI	Heparinase I
HepII	Heparinase II
HepIII	Heparinase III
HepIII	Heparin sulfate lyase/Heparinase III
HGA	Homogalacturonan
HP	Heparin

HS	Heparan sulfate
HSPG	Heparan sulfate proteoglycans
Hyl	Hyaluronate lyase
IdoA	Iduronic acid
KDG	2-keto-3-deoxy-D-gluconate
KS	keratan sulfate
ManA/GulA	Mannuronic acid/Guluronic acid
MBP	Maltose Binding Protein
NMA	Normal Mode Analysis
PG	Polygaluronic
PGA	Polygalacturonan
PGM	Polygaluronic and Polymannuronic
PL	Polysaccharide lyase
PM	Polymannuronic
PolyG	Homopolymeric G block
PolyM	Homopolymeric M block
PolyMG	Alternating M and G monomers
PULs	polysaccharide utilization loci
Rha3S	3-sulfated rhamnose
RG	Rhamnogalacturonans
Rha	L-Rhamnose
RMSD	Root-Mean-Squares Deviation
TEV	Tobacco Etch Virus
TSRI	The Scripps Research Institute

1. Introduction

Living systems are made up of four types of macromolecules: nucleic acids, proteins, lipids and carbohydrates. Carbohydrates are the most complex of these macromolecules and are synthesized and metabolized by all living organisms. They are often present as polymers (either oligo or polysaccharides) linked to proteins or lipids and possess multiple chiral centers. The monosaccharides are connected by α or β glycosidic linkages and exist as linear (e.g. alginate, hyaluronan) or branched polymers (e.g. glycogen). They undergo a wide range of modifications, including acetylation, methylation, oxidation, and sulfation (Bertozzi and Kiessling, 2001; Cummings and Stephen, 2007). The stereochemical variability, linkages, diversity, branching and modifications present in polysaccharides gives rise to their structural complexity (Rodrigues et al., 2011).

The structural complexity of polysaccharides lead to their functional diversity in living organisms. Polysaccharides are known to participate in a wide variety of cellular functions. They function as energy storage molecules (e.g. starch, glycogen) and aid in structural rigidity (chitin, cellulose). Polysaccharide capsules on the bacterial cell surface mediate an interaction between the bacterium and host. This capsule also promotes survival of the bacterium and protects it from immune system assaults after infection (Corbett and Roberts, 2009). In eukaryotes, glycosaminoglycans (GAGs), either free (hyaluronan) or attached to core proteins (proteoglycans) are abundant in the extracellular matrix (ECM) and on the cell surface where they maintain ECM integrity and are involved in a variety of biological processes. Other polysaccharides (e.g. alginate, ulvan, cellulose, pectin and agarose) are predominant in plants, marine bacteria, algae, fungi and are significant components of biomass (Hehemann et al., 2014; Grondin et al., 2017). They have enormous industrial and medical applications. For example, hydrogels derived from cross-linked polysaccharides are employed as drug delivery systems and in tissue engineering (Laurienzo, 2010; Senni et al., 2011).

Two predominant reaction mechanisms are found in enzymes that degrade polysaccharides: hydrolysis and lytic cleavage. Hydrolysis is catalyzed by glycoside hydrolases (GH) that utilizes water to break the glycosidic bond, while polysaccharide lyases (PL) break the glycosidic linkage by abstracting a proton from the C5 position of the sugar. This leads to the

formation of an unsaturated product at the nonreducing end. Extensive information on polysaccharide processing enzymes are available in the CAZy database (http://www.cazy.org) (Lombard et al., 2014). The research described in this thesis is focussed on the determination of the structure of **PL** family enzymes and on characterizing their catalytic mechanism. Using X-ray crystallography, mutagenesis and enzyme activity assays, the structure and catalytic mechanism of four polysaccharide lyases were studied. These are (1) Heparinase III (**HepIII**) from gut bacterium *Bacteroides thetaiotaomicron*, (2) Ulvan lyase **LOR107** from *Alteromonas* species, (3) Ulvan lyase **PLSV3936** from *Pseudoalteromonas* species and (4) Ulvan lyase **NLR48** from *Nonlabens ulvanivorans.* In the following sections, a general background on the mechanisms of polysaccharide degradation is given. This is followed by a detailed overview, specifically on uronate polysaccharides, classification of polysaccharide lyases, their structural diversity and catalytic mechanisms.

2. Literature review

2.1 Polysaccharide degrading enzymes

2.1.1 Glycoside Hydrolases (GHs)

Polysaccharide breakdown catalyzed by GHs is the most common and well understood reaction mechanism. Currently, GHs are classified into ~150 families based on their aminoacid sequences in the CAZy database. These enzymes employ a water molecule to facilitate the cleavage of a glycosidic bond between the anomeric carbon and the bridging oxygen. Hydrolysis of the glycosidic bond can occur with an inversion or retention of configuration of the product's anomeric carbon after cleavage. Both mechanisms have a pair of amino acids with carboxylic groups at the active site and involve oxacarbenium ion-like transition states. (Rye and Withers, 2000).

In the **inverting** hydrolase mechanism, one carboxylate acts as a general base and promotes the nucleophilic addition of a water molecule onto the anomeric carbon while the other carboxylate acts as an acid and donates its proton to the leaving group. This reaction proceeds through a single-displacement mechanism (Rye and Withers, 2000; Vuong and Wilson, 2010) **(Figure 2.1A)**.

In the **retaining** hydrolase mechanism, the reaction occurs through double displacement. In step 1, one of the carboxylates in the active site acts as a nucleophile and attacks the C1 anomeric carbon to form the covalent enzyme-substrate intermediate. Another carboxylate known to function as an acid/base donates a proton to the leaving group. In step 2, the carboxylate that donates a proton to the leaving group acts as a base and coordinates the water for nucleophilic attack on C1, resulting in a cleavage of the enzyme-substrate intermediate (Rye and Withers, 2000; Vuong and Wilson, 2010) **(Figure 2.1B)**.

2.1.2 Polysaccharide Lyases (PLs)

Uronic acid-containing polysaccharides can be degraded by a lytic mechanism. The uronic acid sugar contains a carboxylic group at the C6 position. This depolymerization reaction is dependent on the presence of this carboxylic group and is catalysed by polysaccharide lyases (PL enzymes). Examples of uronate-containing polysaccharides are GAGs, pectin, alginate, ulvan and

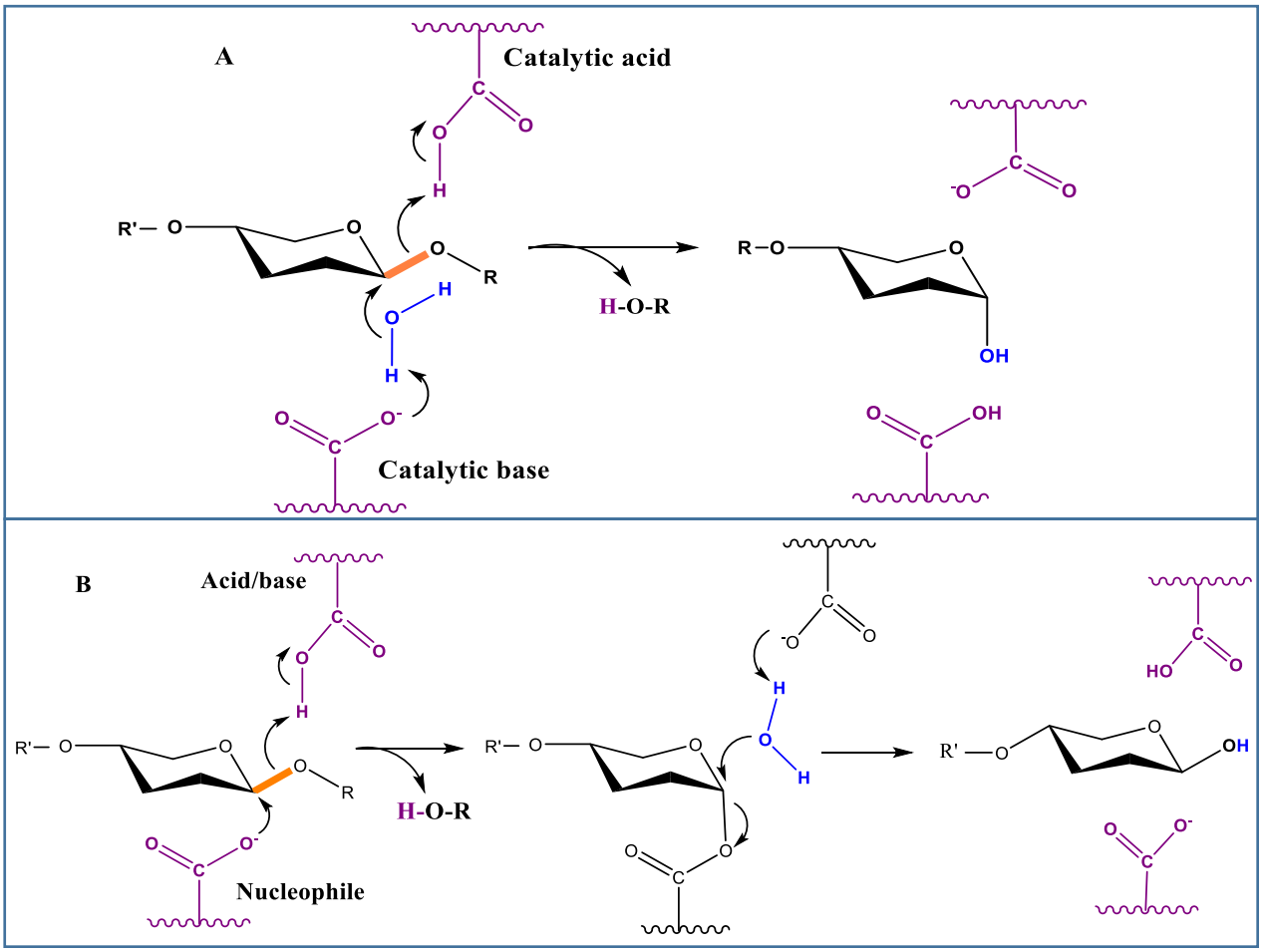

Figure 2:1 Catalytic reaction mechanism of GH.
(A) Schematic of inverting hydrolase mechanism. **(B)** Schematic of retaining hydrolase mechanism (Vuong and Wilson, 2010). All the reaction mechanisms and schematics of polysaccharides in this thesis were drawn using chemdraw.

xanthan. So far, there are 27 sequence related PL families classified in the CAZy database. Many enzymes of bacterial, viral, algae, fungal and plant origin employ lytic mechanisms to depolymerize the uronate-containing polysaccharides. The lytic elimination mechanism occurs in three steps. First, the C6 carboxylate group of the uronate substrate is neutralized to reduce the pKa of the C-5 proton. Second, the C-5 proton is abstracted by a base to form an enolate anion intermediate. Finally, the electrons rearrange to form a double bond between C-4 and C-5 with concomitant cleavage of the C4–O1 glycosidic bond. This degradation mechanism leads to the formation of an unsaturated sugar at the newly formed non-reducing end while the sugar on the newly formed reducing end remains saturated **(Figure 2.2)** (Feingold and Bentley, 1987). Relative to the GH family of enzymes, PLs are somewhat understudied. Thus PL enzymes and their

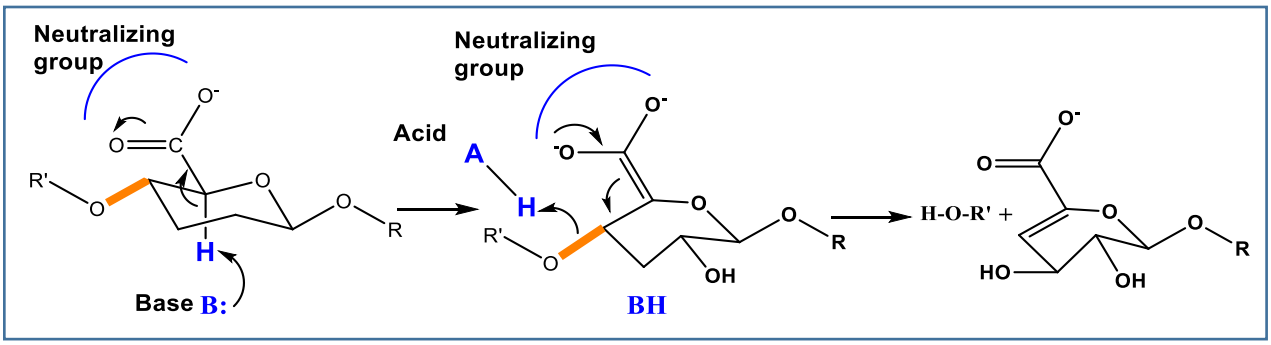

Figure 2:2 Schematic of β-elimination mechanism of PL.

associated lytic mechanisms are the focus of this thesis. The lytic elimination mechanism is described in detail in a later section (section 2.5).

2.1.3 Subsite nomenclature

A common naming pattern of the subsite has been proposed for polysaccharide processing enzymes. The sugars of the substrate are numbered in both directions from the cleaved bond, with a '-' sign on the nonreducing side and a '+' sign on the reducing side of the bond. Thus, the bond to be cleaved is between -1 and +1 sugars (Davies et al., 1997). The residues of the enzyme that interact with individual sugars of the substrate are called the subsites and are given the number of the sugar they interact with **(Figure 2.3)**.

2.2 Uronate-containing polysaccharides

Polysaccharides containing uronic acid are abundant in nature. They are present in algae, bacteria and higher eukaryotes such as plants and vertebrates. These polysaccharides are composed of repeating disaccharide units, containing a six membered sugar linked to a uronate sugar. Based on the nature of the uronic acid sugar at the +1 position, the uronate polysaccharides are classified into three groups: (1) galacturonic acid (GalA) - this group contains pectate and pectin, (2) glucuronic acid (GlcA), iduronic acid (IdoA) - this group includes GAGs, namely hyaluronan, chondroitin, chondroitin sulfate, dermatan sulfate, heparin, heparan sulfate, heparosan, glucuronan, and non-GAGs namely xanthan, ulvan and (3) mannuronic acid (ManA)/guluronic acid (GulA) - this group contains alginates (Garron and Cygler, 2010). All the aforementioned uronates are stereoisomers of one another: GlcA and IdoA are C5 epimers, IdoA and GalA are C4 epimers, IdoA and GulA are C2 epimers, GlcA and ManA are C2 epimers, GulA and ManA are C5 epimers **(Figure 2.4)**.

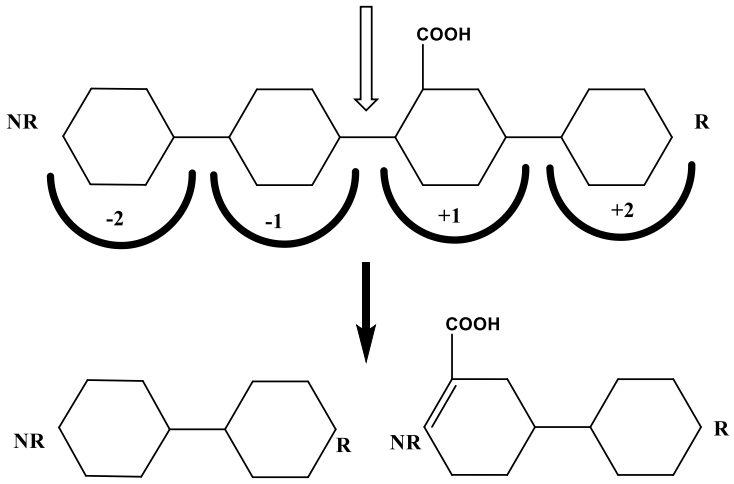

Figure 2:3 Nomenclature of subsites in polysaccharide processing enzymes.
The non-reducing end (NR) of the substrate is on the left and reducing end (R) is on the right. The scissile bond to be cleaved is between +1 and -1 subsite is shown by the arrow.

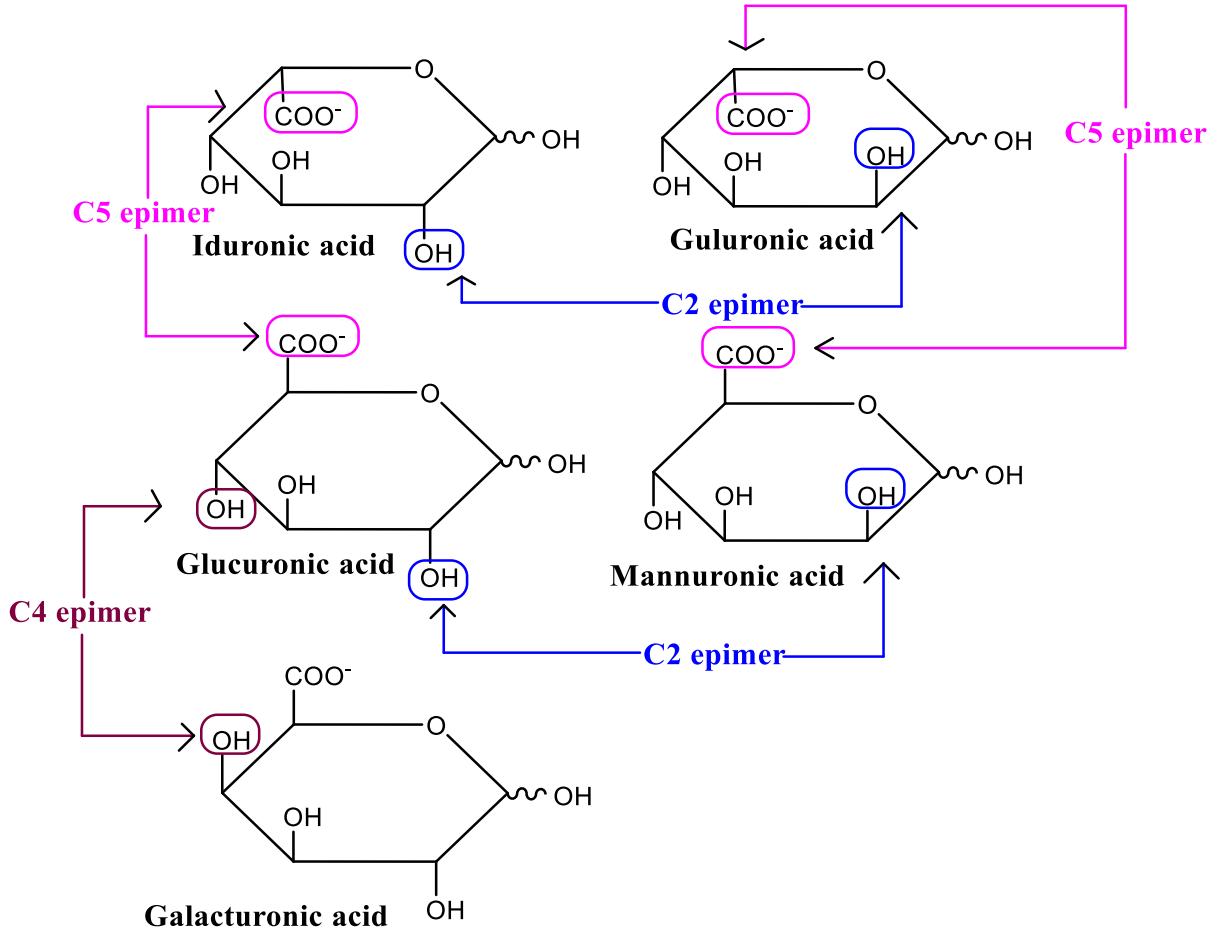

Figure 2:4 Stereochemistry of uronate monomers.

2.2.1 Pectin/Pectate (Galacturonic acid (GalA) containing PL substrate)

Pectin is a complex structural cell wall polysaccharide present in plants. The monosaccharide components of pectin are D-GalA and L-Rhamnose (Rha). Pectins are classified into three types based on the sugar component, degree of modification and branching.

The first type is called homogalacturonan (HGA) or polygalacturonan (PGA) and is composed of repeating α-1, 4-linked D-GalA residues. In PGA, the carboxylate groups are methyl esterified and, to a lesser extent, the C-2 and/or C-3 positions are acetylated. This region of pectin is often called the "smooth" region. The other two types are called Rhamnogalacturonans (RG I, RG II) and have branched pectin with a backbone composed of alternating units of L-Rha and D-GalA. RG1 and RG2 are often called the "hairy" region. In RG1 and RG2, L-Rha undergoes many modifications and the C-2 and C-3 positions of D-GalA are partially acetylated (Hugouvieux-Cotte-Pattat et al., 2014; Abbott and Boraston, 2008; Ridley et al., 2001). Pectin plays an important role in plant growth, morphology, and plant defense mechanism. In addition, pectin has many industrial applications. For example, pectin is used as a gelling and stabilizing polymer in the cosmetic industry. Pectin is also employed in the production of a variety of specialty products including edible, biodegradable films, paper substitutes, and adhesives (Mohnen, 2008).

2.2.2 Glucuronic (GluA) /Iduronic (IduA) containing polysaccharides

The glycosaminoglycans (GAGs), xanthan and ulvan are the uronate polysaccharides in this category. GAGs are a family of linear, sulfated, heterogeneous polysaccharides present in the extracellular matrix (ECM) and on the surface of animal cells. GAGs are composed of repeating disaccharide units. These disaccharides are predominantly made of GlcA or IdoA linked to a hexosamine, either N-acetylglucosamine (GlcNAc) or N-acetylgalactosamine (GalNAc) (Mizumoto and Sugahara, 2012). GAGs are classified into four groups based on differences in the repeating disaccharide units: (1) heparin (HP) / heparan sulfate (HS), (2) chondroitin sulfate (CS) / dermatan sulfate (DS or CS-B), (3) hyaluronan (HA) and (4) keratan sulfate (KS). With the exception of HA, all GAGs undergo post-polymerization modifications on the sugars including sulfation, epimerization and desulfation (De Angelis, 2012). Some bacteria produce polysaccharides with a backbone structure resembling that of GAGs.

2.2.2.1 Heparin (HP)/Heparan sulfate (HS)

Heparin sulfate (HS) consists of repeating disaccharide units of GlcA and GlcNAc. The disaccharide units are linked in this fashion [→4) β-D-GluA (1→4)-α-D-GlcNAc (1→] and are primarily non-sulfated or mono-sulfated (N-sulfated at C2). Heparin (HP) is a form of HS, with a high degree of sulfation and contains IdoA. The major repeating disaccharide sequence of HP is [→4)-α-L-IdoA (1→4)- α-D-GlcNAc) (1→] with sulfation occurring at C2, C6 of GlcNAc and O-sulfonation at C2-IdoA. The difference between HS and HP is in their GlcA/IdoA composition and degree of sulfation. HS has a high GlcA composition and contains ≥1 sulfate group per disaccharide, whereas HP contains high IdoA and ~2.7 sulfate groups per disaccharide, on average (Shriver et al., 2012; Cummings and Pierce, 2010).

Both HS and HP are linked to a serine residue of a number of core proteins and function as proteoglycans (HSPG) (e.g. serglycin, glypican and syndecan). HSPG participates in various cellular functions including blood coagulation, wound healing, embryonic development, and viral infections (Sarrazin et al., 2011). Defects in HSPG biosynthesis are linked to human multiple exotoses syndrome and skeletal developmental abnormalities (Huegel et al., 2013).

2.2.2.2 Chondroitin sulfate (CS)/Dermatan sulfate (DS)

Chondroitin sulfate (CS) is made of repeating units of GlcA and GalNAc with the linkage [→4) β-D-GlcA (1→3) β-D-GalNAc (1→]. Variations in the sulfation pattern of CS give rise to many CS subtypes: (**1**) chondroitin 4-sulfate or chondroitin sulfate A (CS-A) has GlcA (1→3) GalNAc (4S), (**2**) chondroitin 6-sulfate or chondroitin sulfate C (CS-C) has GlcA (1→3) GalNAc (6S), (**3**) chondroitin sulfate D has GlcA (2S) (1→3) GalNAc (6S), (**4**) chondroitin sulfate E has GlcA (1→3) GalNAc (4S and 6S) (Mikami and Kitagawa, 2013; Sugahara et al., 2003).

Dermatan sulfate (DS) has a similar backbone to CS but GlcA is replaced by IdoA and sulfated at the C2 position. DS has the linkage pattern of [→4) β-IdoA (1→3) β-GalNAc (1→] (Trowbridge and Gallo, 2002). DS has a subtype called chondroitin sulfate iE, which has (IdoA) (1→3) GalNAc (4S and 6S) (Sugahara et al., 2003). In addition to their structural roles, CS/DS has an important role in central nervous system (CNS) development, wound repair, infection, growth factor signaling, morphogenesis and cell division (Schwartz and Domowicz, 2002; Mikami and Kitagawa, 2013; Thelin et al., 2013).

2.2.2.3 Hyaluronan (HA)

Hyaluronan (HA) is the only glycosaminoglycan lacking post-polymerization modifications. It functions as a free polysaccharide without being attached to a core protein and is built of repeating disaccharide units comprised of GlcA and GlcNAc with linkage [4→) β-GlcA (1→3) β-GlcNAc (1→] (Gandhi and Mancera, 2008).

2.2.2.4 Keratan sulfate (KS)

Keratan sulfate (KS) consists of repeating disaccharides of Gal and GlcNAc with the linkage [3→) β-Gal(1→4) β-GlcNAc(1→].The post-polymerization modification in keratan sulfate is sulfation at the C6 position of both hexose moieties (Funderburgh, 2002; Garron and Cygler, 2010).

2.2.2.5 Biological role of GAGs

GAGs are abundant in the ECM and connective tissues. They contribute to the native folding and function of proteoglycans, as well as to tissue and organ behavior. In fact, GAGs are involved in stabilization of the fibrillar ECM, control of hydration, regulation of tissue, and organism development by controlling the cell cycle, cell behavior, and differentiation (Lin. 2004, Mizuguchi, *et al*., 2003, Perrimon *et al*., 2000). The negative charges on glycosaminoglycans can bind large quantity of water molecules onto the cartilage and thereby ensures tissue visco-elasticity, which is essential for the proper functioning of the locomotive system.

2.2.2.6 Ulvan (GlcA/IdoA containing polysaccharide from marine algae)

Ulvan is a complex sulfated polysaccharide biosynthesized by the seaweed marine green algae genera *Ulva* and *Enteromorpha*. Ulvan constitutes one of the major cell wall polysaccharides and contributes 8-29% of the dry weight of *ulva* species (Colle et al., 2011). This water-soluble polysaccharide is composed predominantly of 3-sulfated rhamnose (Rha3S), GlcA, IdoA and xylose. The most common repetitive units within the ulvan polysaccharide are disaccharide units [→4)- β-D-GlcA-(1→4)-α-L-Rha3S-(1→] called type A ulvanobiourinic-3-sulfate (A3S) and [→4)-α-L-IdoA-(1→4)-α-L-Rha3S(1→], called type B ulvanobiouronic-3-sulfate (B3S) (Lahaye and Robic, 2007). The presence of IdoA and sulfated rhamnose differentiates ulvan from other polysaccharides of marine origin and displays similarity with mammalian GAGs such as CS and

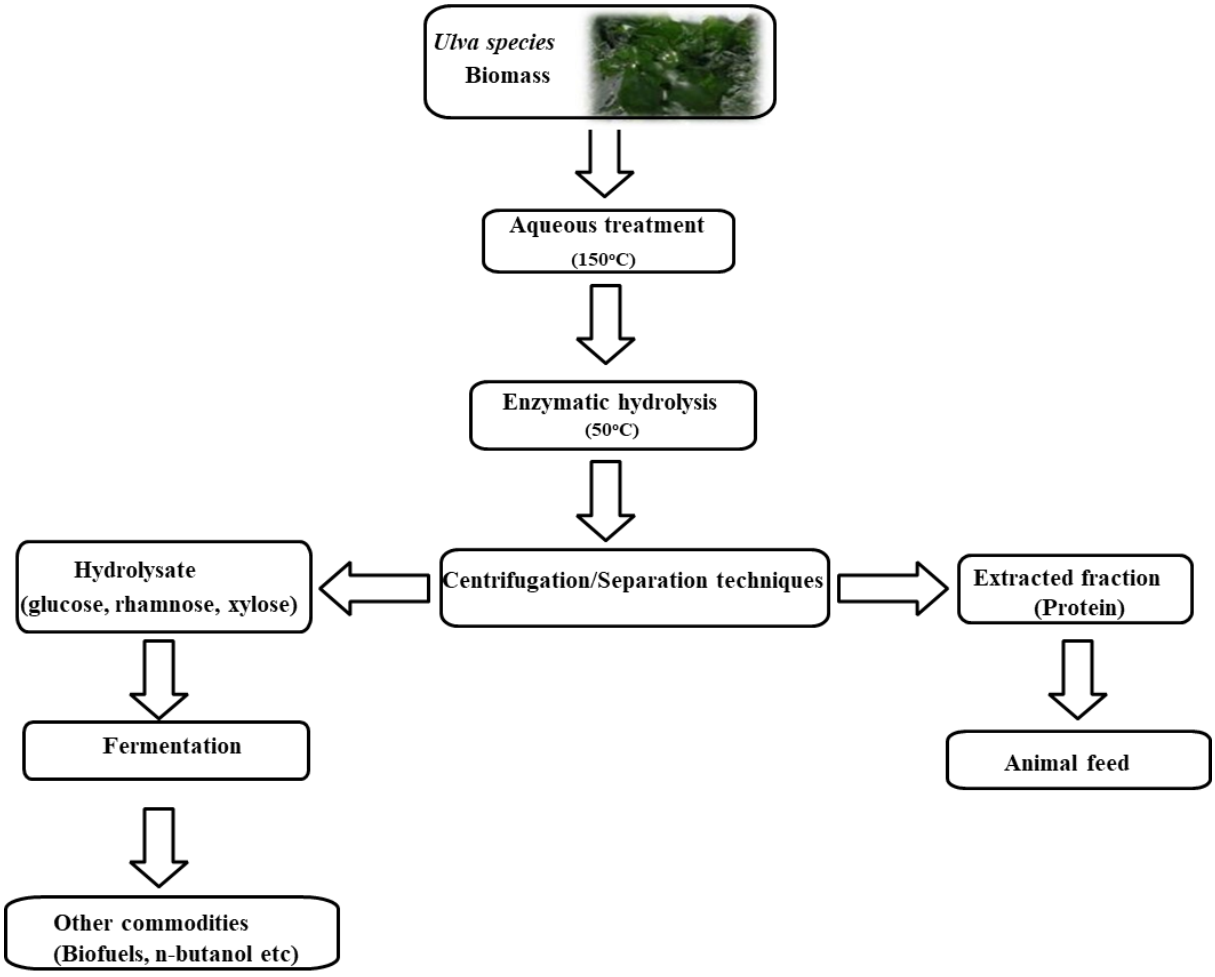

Figure 2:5 Ulvan biomass conversion processes.
Schematic representation of the steps involved in the beneficial degradation of ulvan. Figure adapted and modified from (Bikker et al., 2016).

HA. This distinctive chemical feature makes ulvan an attractive candidate for various biomedical, nanobiotechnological and drug delivery applications. Also, the physicochemical and biological properties of ulvan make it of interest for a variety of industrial applications in the food, agriculture, chemical and biomaterials industries (Cunha and Grenha, 2016; Manivasagan and Oh, 2016; Morelli et al., 2016; Alves et al., 2013). Sulfated ulvan oligosaccharide has been reported as a synthetic elicitor for disease control in plants (Dumas et al., 2010). In addition to medicine and agriculture, the ulvan polysaccharide has captured attention in environmental and bioenergy fields (Singh and Olsen, 2011; Ulaganathan et al., 2018; Ulaganathan et al., 2017).

The green algae called sea lettuce (*Ulva lactuca*) is a common dietary component in Asian and Mediterranean countries. Some green algae are notorious for eutrophication and causing green tides. The expanding environmental impact of these green algae has prompted efforts to mitigate this biomass or to make better use of it. Characterization of *Ulva spp*. shows high crude protein content of up to 44 % dry weight. Biorefining protein and non-protein constituents of seaweed will meet the protein demand to feed increasing animal populations. The sugar waste Rha, GlcA, IdoA will serve as starting material for various synthetics (e.g. Biofuel) (Bikker et al., 2016) **(Figure 2.5)**.

2.2.2.7 Xanthan

Xanthan is an extracellular hetero-polysaccharide synthesized by the pathogenic bacterium, *Xanthomonas campestris*. The main chain polymer is composed of the repeating Glucose (Glu) disaccharide unit with linkage [→4)-β-D-Glu-(1→4)-β-D-Glu-(1→]. The second Glu at C3 has a branch trisaccharide [β-D-Man-(1→4)-β-D-GlcA-(1→2)-β-D-Man-(1→3]. Both mannose units in the branch can be acetylated at the C-6 position and a pyruvate can be added on the C-4 or C-6 position. Xanthan is used as a common food additive and also in food packaging material (Garron and Cygler, 2010; Ahmad et al., 2015). Xanthan is also useful in pharmaceutical and biomedical applications that involve ingestion (Morris and Harding, 2009) .

2.2.2.8 Glucuronan

Glucuronan is a simple homouronic polysaccharide synthesized by bacteria, fungi and algae. Glucuronan has repeating units of GlcA with linkage [→4)-β-D-GlcA-(1→4)-β-D-GlcA-(1→] and undergoes partial acetylation at the C2 or C3 position. This polysaccharide has many applications in food and pharmacological industries (Garron and Cygler, 2010).

2.2.3 Alginates (Mannuronic (ManA) /Guluronic (GulA) polysaccharides)

Alginate is a linear polysaccharide with repeating disaccharide units composed of β-D-ManA and its C-5 epimer α-L-GulA with the linkage [→4) β-D-ManA (1→4) α-L-GulA (1→]. Alginate is classified into three types depending on the distribution pattern of monomers; homopolymeric M block (PolyM) containing only ManA, homopolymeric G block (PolyG) containing only GulA or alternating M and G monomers (PolyMG). The main sources of alginate are brown seaweed. Alginate is present on the cell well and contributes to nearly 40% of the dry

weight of brown algal biomass (Xu et al., 2017). Bacteria belonging to the *Pseudomonas* and *Azotobacter* genera also produce alginates. In *Pseudomonas* species, alginate-mediated biofilm formation helps to avoid phagocytosis (Leid et al., 2005; Pritt et al., 2007). The gelling properties and the ion chelating properties make alginate an attractive candidate with a wide array of applications in industry (e.g. viscosifier, stabilizer, and a gelling agent in the food, printing and biomaterial industries) (Nalamothu et al., 2014).

2.3 Polysaccharide lyases (PLs)

Uronic acid-containing polysaccharides are depolymerized through a lytic mechanism by PL family enzymes. PL enzymes are mostly of bacterial, fungal, algal or plant origin and are highly substrate specific.

2.3.1 PLs in bacterial infection

Pathogenic bacteria interact with host components and modulate their activities to promote pathogenesis. The key components involved are host GAGs and the **PL** produced by pathogens. The process begins with **adhesion** of microbes to the host cell through interactions with GAGs. Heparan sulfate proteoglycans (HSPG) act as host receptors for adhesion of the antibiotic resistant bacteria *Enterococci* onto host colonic epithelial cells (Sava et al., 2009). Adhesion to GAGs has also been demonstrated in a variety of Gram-positive organisms as follows: *Streptococcus pneumoniae* attaches to HP, HS, and CS; S*taphylococci* adhere to CS-A and HS; *Streptococcus pyogenes* recognizes DS, HS, and HP; and *Streptococcus mutans* binds mainly HP (García et al., 2016).

The next step is **invasion -** One invasion strategy employed by microbes is to degrade the ECM. **PL**s are the key players in this mechanism. During bacterial infection, PLs have been demonstrated to depolymerize GAGs. Hyaluronate lyase (Hyl) produced by *S. pneumoniae* during infection, degrades HA to facilitate bacterial invasion (Li et al., 2000). HepP from *Pseudomonas aeruginosa*, a homolog of the heparinase enzyme, showed a significant contribution to the virulence of this pathogenic bacteria. HepP represents a new potential target for antimicrobial agents (Dzvova et al., 2017; Dzvova et al., 2018). Moreover, it has been demonstrated that many bacterial pathogens use the PL degradation products (*i.e*. anionic polysaccharides) to neutralize cationic antimicrobial factors (Jinno and Park, 2015).

Nutrient source - Pathogenic bacteria utilize GAGs as their carbon source. *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Streptococcus pyogenes*, *Streptococcus intermedius*, and *Mycobacterium tuberculosis* expressing hyaluronan lyase (Hyl), have all been shown to use HA as their sole carbon source (Marion et al., 2012).

Pathogenic bacteria produce GAG degrading enzymes as virulence factors to overcome the mobility barrier caused by GAG in the extracellular matrix. Also, GAGs are a rich nutrient source. Thus, bacteria have a dual advantage in degrading the GAG polysaccharide. For this reason, inhibitors of the GAG enzyme may be potential bacteriostatic agents. The bacterial enzymes (PL enzymes) involved in GAG degradation has no mammalian homologs, and thus presents an attractive target for drug development (Jongkees and Withers, 2011).

Recent research interest for the use of PL as **antimicrobial adjuvants** arises from studies on AlgL, an alginate lyase from *Pseudomonas aeruginosa*. *In vitro* studies have demonstrated that addition of AlgL affects biofilm formation and causes lysis of *P. aeruginosa*. (Thallinger et al., 2013; Alkawash et al., 2006).

2.3.2 PL in Gut Microbiome

Other than pathogenic bacteria, commensal gut-residing symbiotic (friendly) bacteria encode a repertoire of PL enzymes. The gut microbiome community influences a variety of cellular functions. Abnormalities in their composition are associated with inflammatory bowel disease, colon cancer, antibiotic-associated colitis and obesity (Koropatkin et al., 2012). Availability, distribution and composition of the gut microbiome depends on its access to nutrients. Some members (e.g. *Bacteroidetes* species) depend on the host GAGs as a high priority carbon source (Flint et al., 2012; Cartmell et al., 2017). Other symbiotic bacteria feed on the complex dietary polysaccharides from plants and animals and aid in their saccharification as an energy source for the host (Ndeh and Gilbert, 2018; Ndeh et al., 2017). The glycoside hydrolases and polysaccharide lyases required to cleave the complex glycans are not present in the human genome and the microbiome supplements the host with these useful enzymes (Kau et al., 2011).

2.3.3 Plant lyases

Several PL enzymes have been reported as virulence factors in plant pathogens. Targeting the plant cell wall is a simple and efficient pathogenic attack mechanism. Plant pathogenic bacteria and fungi produce an arsenal of enzymes to degrade plant cell wall polysaccharides. Soft rot

infection in plants, caused by pectinolytic microorganisms such as *Erwinia carotovora* and *Erwinia chrysanthemi*, leads to plant tissue maceration and cell death (Hugouvieux-Cotte-Pattat et al., 2014; Abbott and Boraston, 2008).

Uronate-containing polysaccharides represent a substantial biomass (e.g. alginate present in the cell wall of brown algae and pectin polysaccharide of the plant cell wall). Employing polysaccharide lyases with broad substrate specificity is one approach to efficiently degrade this biomass.

Scarcity of fossil fuel resources, steady increase in energy demand, and concern over climate change by greenhouse gas emissions has increased the need to find alternative sources for biofuel production. Microbial fermentation of available biomass is a feasible alternative. In this aspect, cell wall polysaccharides of plants (pectin) and algae (alginate, ulvan) represent a usable biomass (Singh and Olsen, 2011; Medipally et al., 2015). The major drawbacks to use plant polysaccharides for producing fuel are: (1) that they represent a potential food source and (2) the presence of lignin, which is recalcitrant to microbial fermentation. Meanwhile, macroalgae (alginate, ulvan) is an abundant alternative and the absence of lignin in algae makes the extraction simple and inexpensive. The ultimate goal is to breakdown this complex material, which can then be converted into ethanol or other useful compounds. Polysaccharide degrading enzymes and complementary enzymes encoded in the polysaccharide utilization loci (PULs) form a powerful toolkit to aid in this process (Hemsworth et al., 2016). Recent research on an engineered microbial platform showed a simultaneous assimilation and degradation of alginate from brown algae into ethanol. Catabolism of alginate by the engineered microbe produces the common metabolic intermediate pyruvate, from which a plethora of fuels and chemical compounds can be synthesized **(Figure 2.6)**. (Wargacki et al., 2012)

Polysaccharide lyases are also useful in synthetic biology to produce polysaccharides of specific length.

Thus, identifying new PL enzymes has potential applications in a wide array of fields. Studying the structure, catalytic mechanism and substrate specificity of the PL enzymes has been a long standing interest in glycobiology.

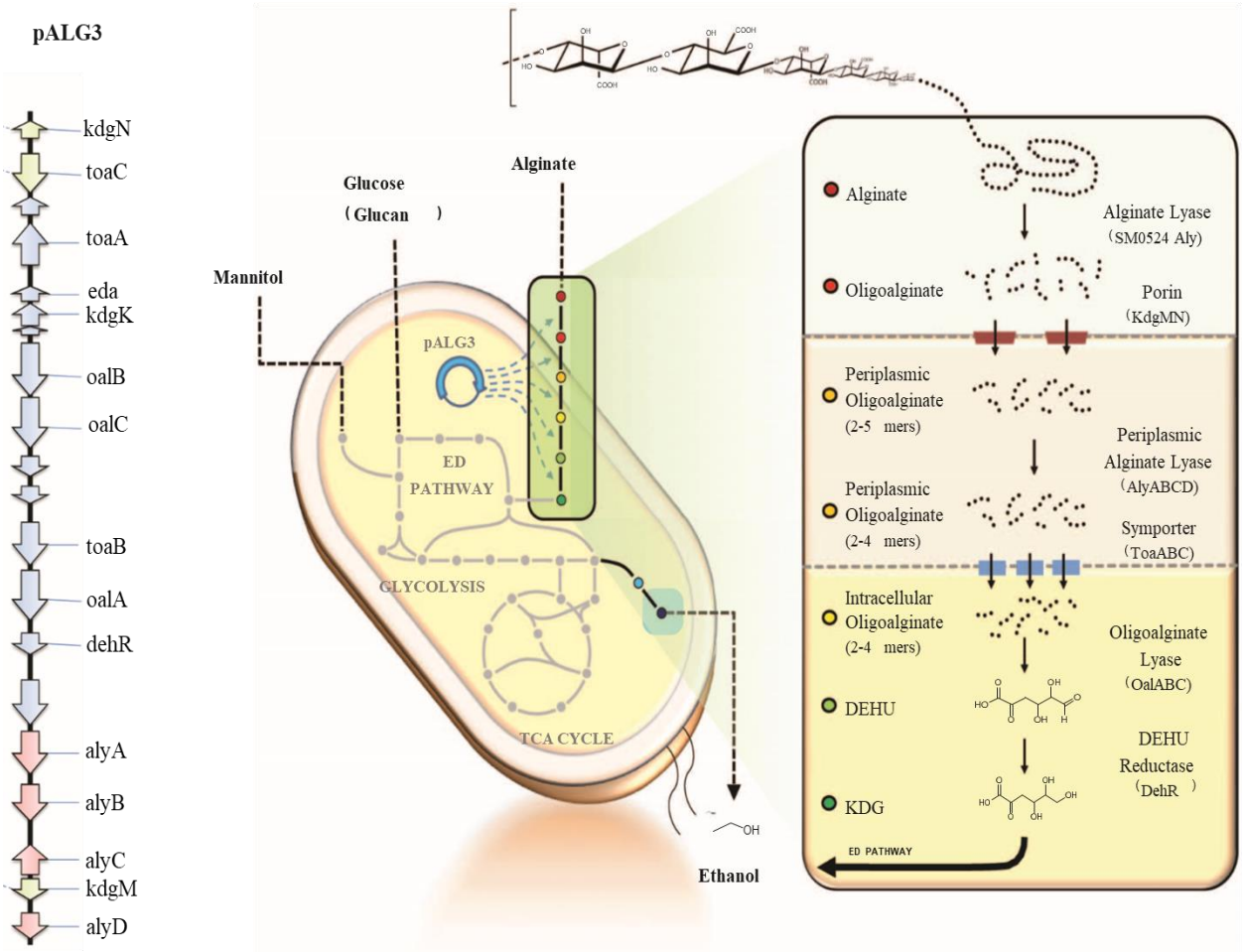

Figure 2:6 Engineered microbial platform to produce biofuel, ethanol other starting material from alginate.

(A) The middle figure shows the schematic representation of *E. coli* engineered for alginate uptake and degradation. pALG3 contains the genes for alginate uptake and metabolism. The different genes in the pALG3 are shown in the left figure. The right figure is the enlarge view describing the steps in alginate metabolism. In the first step, alginate polymers are degraded into oligomers by an alginate lyase followed by transport through the outer membrane porins (KdgMN) into the periplasmic space. In the periplasm, the oligomers are further degraded into di-, tri-, and tetramers by periplasmic alginate lyases (AlyABCD). This is followed by transport into the cytosol through oligoalginate transporters (ToaABC). In the cytoplasm, oligoalginate lyases (OalABC) further degrades the oligomers into monomer units of 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH). DEH reductase (DehR) catalyze the conversion of DEH to 2-keto-3-deoxy-D-gluconate (KDG). The KDG enters the Entner–Doudoroff (ED) pathway. This platform can be employed to produce fuels and chemicals such as ethanol. Figure adapted and modified from (Wargacki et al., 2012)

2.4 Classification of PLs

With the continued efforts in genome sequencing, the number of PL families are increasing over time. There are currently 27 PL families grouped by sequence similarities in the CAZy database. At least one representative structure from each family is now available and representative structures for three such families are described in this thesis. The following section summarizes the known details of fold, substrate specificity and catalytic mechanism of the PL family enzymes, which are **grouped by their substrates**.

2.4.1 GAG lyases

Families classified as GAG lyases depolymerize GlcA/IdoA-containing GAGs. Based on their sequences, they are classified into PL 6, 8, 12, 13, 16, 21 and 23 families in the CAZy database. **PL6** contains **chondroitinase B** (ChonB, EC 4.2.2.19), **PL8** contains GAG lyase with broad substrate specificity, **hyaluronate lyase** (HyL, E.C. 4.2.2.1), **chondroitin AC lyase** (ChonAC, E.C. 4.2.2.5) and **chondroitinase ABC** lyase (ChonABC, E.C. 4.2.2.20), **PL12** contains heparin sulfate lyase/heparinase III (HepIII, E.C. 4.2.2.8), **PL13** contains Heparinase I (HepI, E.C. 4.2.2.7), **PL16** contains **hyaluronan lyase** (HyL, E.C. 4.2.2.1), **PL21** contains Heparinase II (HepII, E.C. 4.2.2.8), acharan-sulfate lyase (EC 4.2.2.-) and **PL23** contains chondroitin lyase (CsL, EC 4.2.2.-). Xanthan lyase and ulvan lyase are exceptions in this category; they degrade GlcA/IdoA containing substrates other than GAGs.

Structure diversity in GAG lyase – Figure 2.7 displays the various structural folds observed in GAG lyase.

2.4.1.1 Multidomain $(\alpha/\alpha)_n$ toroid fold

In general, the $(\alpha/\alpha)_n$ toroid fold contains predominantly two domains; one domain with α-helices and another domain with β strands. The α- helices form helical hairpins and fold into a horseshoe-like toroid shape, where n represents the number of helical hairpins. The number of helical hairpins vary among different families. The β strands fold into antiparallel β sheets.

This is the fold seen in many GAG lyase families (PL8, PL12, PL21, and PL23). HyL, Chon AC and Chon ABC are classified in the **PL8** family and share significant sequence, structure and mechanistic homology (Garron and Cygler, 2010). This lyase subfamily has broad substrate specificity.

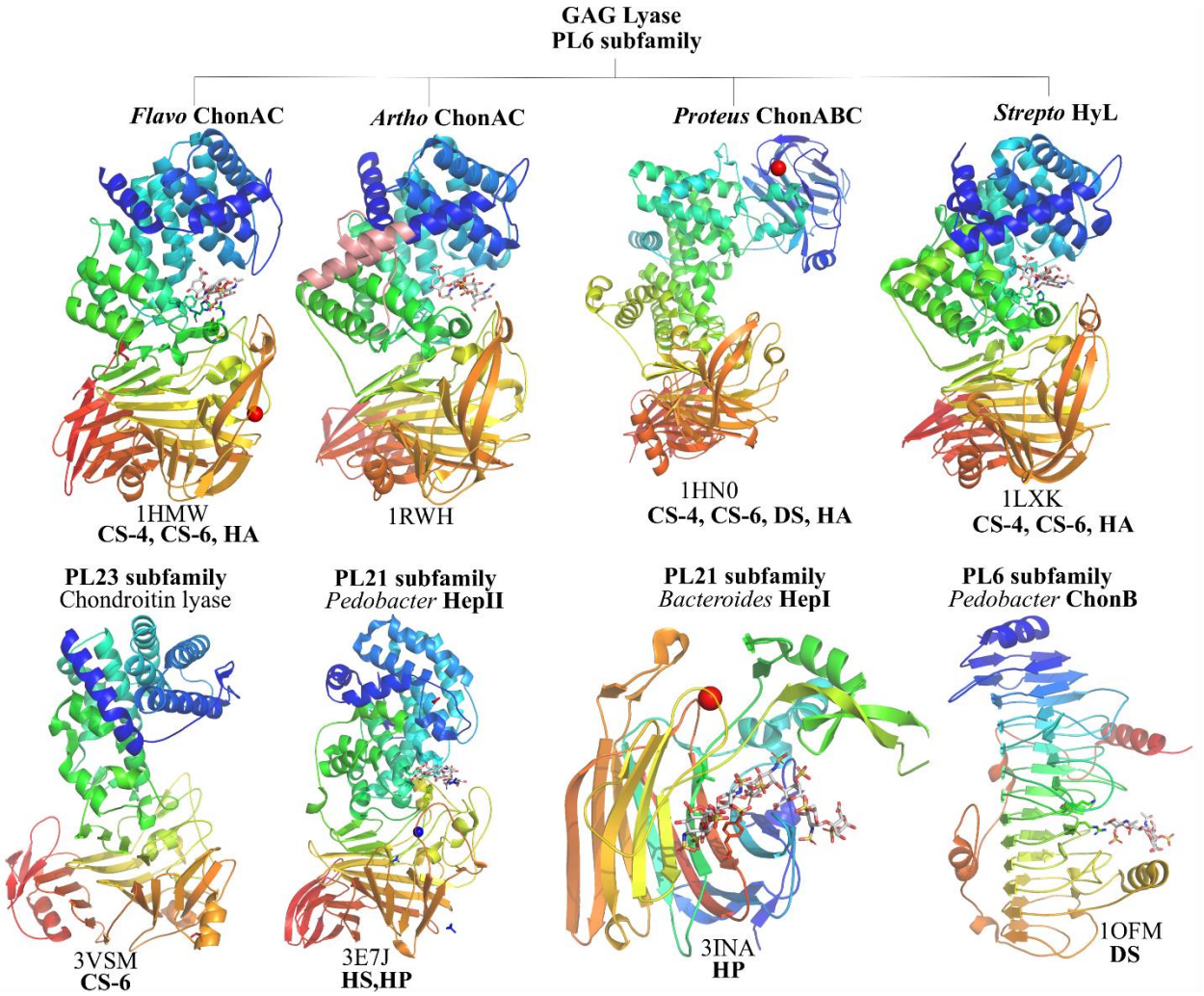

Figure 2:7 Structural Diversity in GAG lyases.

Cartoon diagram of representative structures of GAG lyase with different folds. Their PDB code and the names are shown below each structure. The structures are rainbow colored from blue at the N-terminus to red at the C-terminus. The structure of the enzyme substrate complex is shown for families where complex structures are available. The substrates are shown in stick representation. The top panel shows multi domain folds seen in the PL6 subfamily (ChonAC, ChonABC, Hyl).The bottom panel shows the multi domain fold seen in PL23 (CS lyase), PL21 (HepI) family, the β-jelly roll fold seen in PL21(HepI) and the right handed β-helix fold seen in PL6 ChonB. This figure and the other protein structure figures were made using PyMol (pymol.org).

Chondroitin AC lyase (**ChonAC**) - Crystal structures of ChonAC from *Flavobacterium heparinum* (PDB code: 1HMW) and *Arthrobacter aurescens* (PDB code: 1RWH) are available in the Protein Data Bank (PDB). Both ChonAC structures have two domains with an $(\alpha/\alpha)_5$ toroid fold at the N-terminus and 4 antiparallel β-sheets at the C-terminus. ChonAC has depolymerization activity on CS-A, CS-B and hyaluronan (Rye and Withers, 2002). The two enzymes differ in their

mode of action. *Flavobacterium* ChonAC is an endo-lytic enzyme, whereas *Arthrobacter* ChonAC acts in exo-lytic fashion on the GAG substrate. The difference in the mode of cleavage is attributed to the insertion (~35 residues forming secondary structure of α helices and a loop) present at the N-terminal domain of *Arthrobacter* ChonAC. This minimizes the size of the substrate binding cavity and prevents the binding of large oligosaccharides. The catalytic residues in *Flavobacterium* ChonAC (His225, Tyr234, Arg288, and Glu371) and *Arthrobacter* ChonAC (Asn183, Tyr242, His233) are structurally conserved. (Lunin et al., 2004; Huang et al., 2001; Rani et al., 2017)

Chondroitinase ABC lyase (**ChonABC**) - Crystal structures of ChonABC from *Proteus vulgaris* (PDB code: 1HN0) and *Bacteroides thetaiotaomicron* (PDB code: 2Q1F) have been determined. Despite little sequence identity, the two structures adopt the same fold. Unlike ChonAC, ChonABC has an extra domain at the N-terminus. This domain has a two-layered β-sheet fold resembling the carbohydrate binding domain in cellulase (Huang et al., 2003). The middle domain has a $(\alpha/\alpha)_5$ toroid fold and shows structural homology to ChonAC. The C-terminal domain has four antiparallel β-sheets. Chondroitinase ABC is an endo-lytic enzyme and has demonstrated depolymerization activity on CS-A, DS and HA (Shaya et al., 2008; Huang et al., 2003). Interestingly, the structural determinants for specificity toward DS (the GAG with IduA-C5 epimer of GlcA) have been gleaned from the crystal structure of *Bacteroides thetaiotaomicron* ChonABC (PDB code: 2Q1F). A second cluster of conserved residues were identified 12Å away from the catalytic tetrad. Site directed mutagenesis and enzyme activity assays on the second cluster revealed that His345 is necessary for the breakdown of DS but not for CS (Shaya et al., 2008; Rani et al., 2017).

Hyaluronan lyase (**HyL**) - Crystal structures of HyL from *Streptococcus pneumoniae* and *Streptococcus agalactiae* have been determined in apo and complex form with different GAG substrates. HyL is an endo-lytic, processive enzyme reported to have four domains: an N-terminal domain for GAG binding, a linker domain (which separates the binding and catalytic domain and assists in processivity of the enzyme), a catalytic domain and a C-terminal domain (Rigden et al., 2006). The catalytic domain has an $(\alpha/\alpha)_5$ toroid fold and the C terminal domain has three stacks of antiparallel β sheets (Rigden and Jedrzejas, 2003). The overall structure resembles other PL8 family members ChonAC and ChonABC. One interesting feature learnt from several crystal structures of HyL is its flexibility and domain movement. The observed domain movements open

or close the substrate binding site. Indeed twisting and bending of the domains has been related to the processivity and catalytic mechanism of the enzyme (Rigden et al., 2006).

PL23 family contains chondroitin lyase (**CsL**) (EC 4.2.2.) (ODV-E66) from *Baculovirus*. Even though they share <12% sequence identity, the structural features are similar to the PL8 family of lyases. The overall fold is α/α toroid with antiparallel β sheets. The CsL is reported to have specificity for non-sulfated or the CS-C substrate (6S sulfated) (Kawaguchi et al., 2013).

PL21 – Heparinase II **(HepII**) is classified in the PL21 family and acts endolytically, cleaving both heparin sulfate and heparin GAGs. The crystal structure of hepII from the soil bacterium *Pedobacter heparinus* showed that the enzyme forms a homodimer with two independent active sites in each monomer. The overall fold is $(\alpha/\alpha)_n$ toroid with antiparallel β sheet and composed of three domains: An N-terminal domain forms an $(\alpha/\alpha)_6$ helical hairpin toroid, a central domain contains two layered β sheets and the C-terminal domain containing β-strands is folded like a β- barrel. HepII shows overall structural similarities to the PL8 HyL, ChonAC and PL15 alginate lyase. A Zn^{2+} ion is reported at the interface of the N-terminal and the central domain and plays a structural role. The catalytically important residues deduced from the complex structures are His406 and Tyr257 for both GlcA and IdoA containing HS and HP. Meanwhile, similar to ChonABC, a distant Histidine residue His202, is necessary for the degradation of IdoA (C5 epimer of GlcA) containing HP (Shaya et al., 2006; Shaya et al., 2010).

PL12 - Heparinase III (**HepIII**) from *Bacteroides thetaiotaomicron* is classified in the PL12 family. Heparinase enzymes (HepI/II/III) of bacterial origin were first isolated from the soil bacteria *Pedobacter heparinus* and have been used to study the structure and sulfation patterns in HP and HS (Desai et al., 1993). Structural studies of heparinase enzymes HepI (PDB code: 3ILR) (Han et al., 2009) and HepII (PDB code: 2FUQ, 3E7J) from *P. heparinus* (Shaya et al., 2010) have been previously carried out in our laboratory and their catalytic mechanisms were proposed. Heparinase enzymes show no recognizable sequence similarities. Also, HepIII has a different substrate specificity than HepI/II and prefers low sulfated, GlcA rich regions of HS. Efforts to determine the structure of HepIII from *P. heparinus* were unsuccessful. To overcome this obstacle, an orthologous protein BT4657 from *Bacteroides thetaiotaomicron* was cloned in our laboratory. The first part of this thesis research will describe the structure and catalytic mechanism of the BT4657 gene product.

2.4.1.2 β-jelly roll fold

The common feature of the β-jelly roll fold has two antiparallel β-sheets (SA, SB) stacking against each other. They are bent nearly 90° in the middle, forming a substrate binding groove that extends the length of the enzyme. The inner sheets (SA) of the groove have the conserved residues which perform the depolymerization reaction. The number of strands in the inner and outer sheet varies among the different PL families. The loops lining the β-sheets are of variable lengths. A disulfide bridge is seen in a few PL enzymes with this fold (Garron and Cygler, 2010).

PL13- Heparinase I (**HepI**) (PDB code: 3INA) from *Bacteroides thetaiotaomicron* is well characterized and structures of the enzyme in complex with varying substrate lengths have been reported. HepI acts in endo-lytic fashion and cleaves the HP substrate. HepI has the classical β-jelly roll fold, two antiparallel β-sheets stacking together with a thumb-like insertion domain. The structural integrity of HepI is maintained by a Ca^{2+} ion at the hinge region between the β-jelly roll fold and the thumb domain. This thumb domain facilitates binding of the heparin substrate, because deleting the thumb domain impairs the substrate binding of HepI (Han et al., 2009).

2.4.1.3 Right-handed β-helix fold

This is the fold predominantly seen in extracellular pectate lyase from plant pathogenic bacteria or fungi. The **ChonB** lyase is classified within the **PL6** family and adopts an unusual right-handed β-helical fold. The architecture of the β-helix fold comprises three β-sheets PB1, PB2, and PB3. The sheets are connected by turns or loops, designated T1 (connecting PB1 and PB2), T2 (connecting PB2 and PB3) and T3 (connecting PB3 and PB1) (Huang et al., 1999). This fold is well adapted for the recognition of highly polyanionic polysaccharides. ChonB from *Pedobacter heparinus* specifically degrades DS (IdoA) in an endo-lytic manner. The β-helical fold in ChonB has 13 coils and possesses all of the common stacking interactions (aromatic stack, aliphatic stack and asparagine ladder) observed in the β-helical fold. The interactions are extensive and involve more residues than other proteins having this fold (Michel et al., 2004). PL6 also contains 3 alginate lyases and ChonB shows significant sequence homology to polyM specific alginate lyase AlyP (398 amino acid residues) from *Pseudomonas* sp. OS-ALG-9 (Xu et al., 2017). ChonB differs from ChonAC and ChonABC lyases not only in its fold but also in the mode of substrate recognition and the catalytic mechanism. The former interacts with substrates mostly through ionic interactions whereas the latter utilizes aromatic stacking (CH/π bonds interactions)

for substrates binding. This could explain the broad substrate specificity of ChonAC facilitating the binding of GAGs with varying sulfation patterns. In ChonB, the Ca^{2+} ion in the active site plays a catalytic role; Lys250 and Arg 257 are also reported to be essential for catalysis (Michel et al., 2004; Garron and Cygler, 2010; Rani et al., 2017).

2.4.1.4 Triple Strand β-helix

PL16 - Several crystal structures of **HyLs** from *Streptococcus pyogenes* have been determined. They are composed of three intertwined polypeptides, which fold independently into three β-helices (Garron and Cygler, 2010). PL16 HyLs are different from PL8 HyLs in both fold and substrate specificity. PL16 HyLs degrades only HA and not CS. The overall architecture is elongated with an α/β capping at the N-terminal domain, which is connected to the central right-handed β-helix domain by two helical coils. The central domain resembles an irregular triangular tube and the β-strands of the domain are bent, with their concave surface facing outside the helix. This creates a depression along each side of β-helix, which serves as the substrate binding site for HA polysaccharide chains (Smith et al., 2005; Mishra et al., 2009).

2.4.2 Alginate lyases

Alginate polysaccharides are degraded through a β-elimination mechanism by PL enzymes. Based on sequence similarity, these enzymes are classified into seven different families (PL5, 6, 7, 14, 15, 17 and 18) in the CAZy database. Alginate lyases are identified in both prokaryotes and eukaryotes as well as in viruses. So far, the structure of at least one alginate lyase is available for each of these 7 families. The common structural folds seen in alginate lyase are; β-jelly roll, $(\alpha/\alpha)_n$ toroid and β-helix fold. Both endo-lytic and exo-lytic alginate lyases with broad and narrow substrate specificities have been observed.

The following section describes the structural diversity in alginate lyases and **Figure 2.8** displays the various folds seen in alginate lyase.

2.4.2.1 β-jelly roll fold

This fold is seen in PL7, PL13, PL14, PL18 and PL20. The well-studied alginate lyase from the **PL7** family is **A1-II'** from *Sphingomonas* (PDB code: 2CWS). The complex structure with trisaccharide and tetrasaccharide substrates have been reported for this alginate lyase. A1-II' is an endo-lytic alginate lyase with broad substrate specificity acting on PolyG, PolyM and

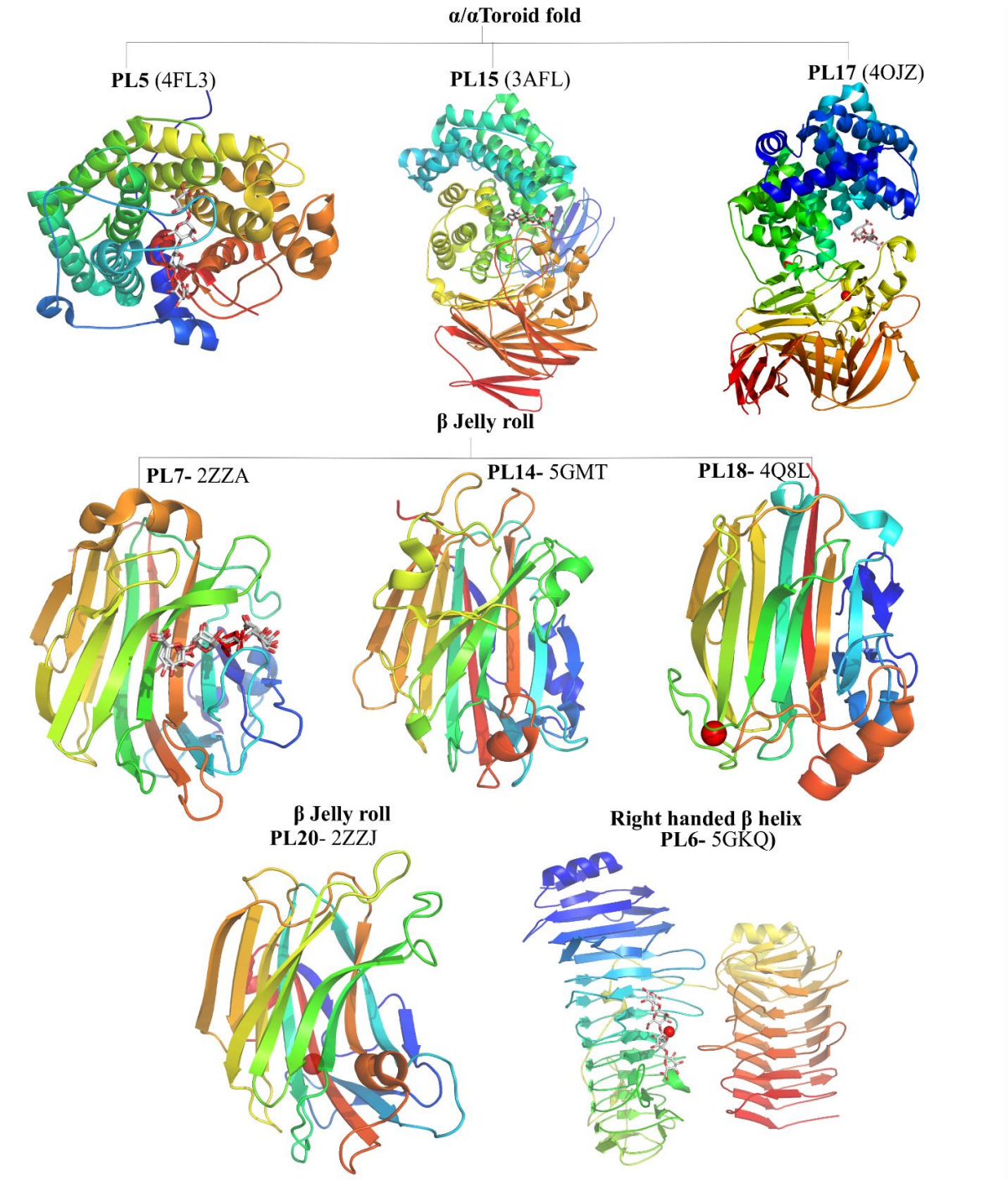

Figure 2:8 Structural diversity of alginate lyases.

Cartoon diagram of representative structures of alginate lyase with different folds. Their PDB code and the names are shown at the top of each structure. The structures are rainbow colored from blue at the N-terminus to red at the C-terminus. The structure of the enzyme substrate complex is shown for families where complex structures are available. The substrates are shown in stick representation. The top panel shows the single and multi-domain (α/α) toroid fold seen in PL5, 15 and 17. The middle panel and the bottom left panel shows the β-jelly roll fold seen in PL7, 14 and 18. The bottom right panel shows the right handed β-helix fold seen in PL6.

PolyMG (Ogura et al., 2008). The other enzymes classified in PL7, namely **PA1167** (PDB code: 1VAV) from *Pseudomonas aeruginosa* (Yamasaki et al., 2004), **AlyA** (PDB code: 4OZX) from *Klebsiella pneumoniae*, **AlyPG** (PDB code: 1UAI) from *Corynebacterium* (Osawa et al., 2005) and **AlyA1** (PDB code: 3ZPY) from *Zobellia galactanivorans* are also endo-lytic, with the canonical elongated active site feature, whereas **AlyA5** (PDB code: 4BE3) from *Zobellia galactanivorans* is an exo-acting lyase. The active site of AlyA5 has a pocket-like architecture, with one end of the active site groove closed by a long loop (Thomas et al., 2013). **PL14** is the only alginate lyase family with representatives from bacteria, eukaryotes and viruses. The alginate lyase vAL-1 from *Paramecium bursaria Chlorella* shows two pH dependent modes of action; either endo- or exo- acting. The catalytic domain of vAL-1 acts as endo-lytic lyase at neutral pH (releases hexasaccharide) and an exo-lytic lyase under basic conditions (releases disaccharides at pH 10). The ionic charge at the active site is reported to regulate the exo-/endo- action in vAL-1 (Ogura et al., 2009). Alginate lyases classified under **PL18** are bifunctional and endo-lytic. The overall fold and catalytic site architecture are similar to the PL7 family but the presence of a unique N-terminal region in the PL18 family acts as an intramolecular chaperone and facilitates protein folding (Dong et al., 2014). Also, PL18 family enzymes have calcium ions, which play a central role in orienting the catalytic residues. Such Ca^{2+} ions have not been reported in PL7 enzymes. The **PL20** family has enzymes from both bacteria and eukaryotes. Structural details of 1,4-glucuronan lyase from *Trichoderma reesei* are available (PDB code: 2ZZJ). Like PL18 alginate lyase and PL13 HepI, the 1,4-glucuronan lyase has a Ca^{2+} ion contributing to the stability of the enzyme (Konno et al., 2009).

2.4.2.2 $(\alpha/\alpha)_n$ toroid fold in alginate lyase

This fold is present in alginate lyase families PL 5, 15 and 17. PL5 enzymes have a single domain whereas PL15 and PL17 contain multiple domains. A well-studied **PL5** family member is **A1-III**, a mannuronate specific alginate lyase, from *Sphingomonas* species A1. A1-III has six inner and five outer helices forming a toroid fold with four helical hairpins $(\alpha/\alpha)_4$ (Yoon et al., 1999). Most characterized PL5 alginate lyases are endo-lytic and polyM-specific enzymes from bacteria. But Smlt1473 from *Stenotrophomonas maltophiliacan* shows broad substrate specificity. It utilizes pH as a mechanism to achieve substrate selectivity. At pH 5 it degrades HA substrates, at pH 7 it degrades poly GlcA substrates and at pH 9 it degrades alginate (MacDonald and Berger, 2014).

The enzymes reported in **PL15** and **PL17** are exo-lytic with a multi domain architecture. PL15 family Atu3025 (PDB code: 3A0O) from *Agrobacterium tumefaciens* is a multi-domain alginate lyase with a small N-terminal beta sheet domain, a central toroid fold and a C-terminal β-sheet domain. The active site is between the central toroid fold and the C-terminal domain. The exo-lytic mode of action in PL15 family is related to the conformational changes caused by lid loop motion upon substrate binding (Ochiai et al., 2010). PL17 family Alg17c from *Saccharophagus degradans* is a two-domain protein with an N-terminal toroid fold $(\alpha/\alpha)_6$ and a C-terminal anti-parallel β-sheet domain. As in PL15, the active site is located between the barrel and the C-terminal domain. But PL17 exist as a homodimer and the dimerization results from the insertion of a β-turn from the β-sheet domain to the N-terminal domain of another monomer (Park et al., 2014).

2.4.2.3 β-helix class

Among alginate lyases this fold is only seen in enzymes classified under the PL6 family. The structure of AlyGC from the marine bacteria *Glaciecola chathamensis* was recently solved and the structural features and catalytic mechanism was revealed. The enzymes are exo-lytic and showed preference for a PolyG block. The architecture of the β-helix fold comprises three β-sheets PB1, PB2, and PB3. The sheets are connected by turns or loops T1 (connecting PB1 & PB2), T2 (connecting PB2 & PB3) and T3 (connecting PB3 & PB1). This alginate lyase has two domains; both the N- and C-terminal domains adopt a right-handed β-helix fold. The enzyme exists as a homodimer in solution. Most of the conserved residues are identified in the T3-PB1-T1. The active site is present at the N-terminal domain but the C-terminal domain is also necessary for catalysis and dimerization. This is evident from the loss of enzyme activity and dimerization in a mutant lacking the C-terminus. (Xu et al., 2017).

2.4.3 Pectate lyases

Pectate lyases (E.C. 4.2.2.2) were identified in the genomes of pathogenic bacteria, fungi associated with plants, the gut residing bacterial community, human pathogenic bacteria, fungi, archaea and eukaryotes. These organisms employ a series of enzymes to degrade the plant cell wall in order to obtain their carbon source (Yadav et al., 2009; Herron et al., 2000). Humans do not produce enzymes to degrade pectin, so we depend on the gut microbiome to degrade pectin from plant dietary sources (Flint et al., 2012). In all these organisms, pectin degradation is carried out by a series of pectinases (Hugouvieux-Cotte-Pattat et al., 2014; Ndeh et al., 2017). To

accomplish pectin degradation, some of the pectate lyases are secreted into the extracellular environment. Pectate lyases have also been reported as both exo- and endo-acting enzymes. Most of the characterized endo- pectate lyases act in the extracellular environment and in the periplasmic space and break long pectin polysaccharides into small oligosaccharides of differing sizes. These oligogalacturonans are transported into the cytoplasm where the exo-pectate lyase carry out the degradation process (Abbott and Boraston, 2008; Gilbert, 2010).

Of the 27 families in the CAZy database, members of PL1, PL2, PL3, PL9 and PL10 have demonstrated activities on pectic polysaccharides (Seyedarabi et al., 2010). Members of the PL22 family act on oligogalacturonans, whereas PL4 and PL11 family members degrade rhamnogalacturonans (RG). Structural characterization of these PL families revealed a number of different folds including a right handed β-helix, $(\alpha/\alpha)_n$ toroid fold, β-propeller fold, β-sandwich fold and different requirements for metal co-factors. A brief outline of the structural features, and catalytic mechanism of pectin/pectate lyase follows.

The following section describes the structural features of pectate lyase and **Figure 2.9** displays the various structural folds observed in pectate lyase.

2.4.3.1 Right handed β-helix

The **right-handed β-helix** is the most common fold observed in pectate lyase families. Pectate lyase, belonging to families PL1, PL3, and PL9 share this fold. The right handed β- helix fold was first discovered in PL1 family PelC from *Dickeya dadantii* (M D Yoder et al., 1993). As mentioned in the GAG lyase and alginate lyase section, the core of this fold contains three parallel β sheets forming a helical architecture. Each coil of this helical architecture is composed of three consecutive β strands (PB1, PB2 and PB3) connected by turns (T1, T2 and T3). The stability of the cylindrical core is strengthened by the characteristic "asparagine ladder" and interaction of aliphatic, aromatic side chains in the interior of the core (Yoder et al., 1993; Jenkins et al., 2004). This structural rigidity helps with the secretion of pectinases into harsh and competitive environments, such as the gastrointestinal tract of animals, soil, and plant cell walls (Abbott and Boraston, 2008). Sequence variability in the PB1 strands and T3 turns reflect adaptations in the substrate binding site and catalytic region to accommodate specific substrates (Garron and Cygler, 2010). The **PL1** family has enzymes from all classes of living organisms. Most notably, PL1

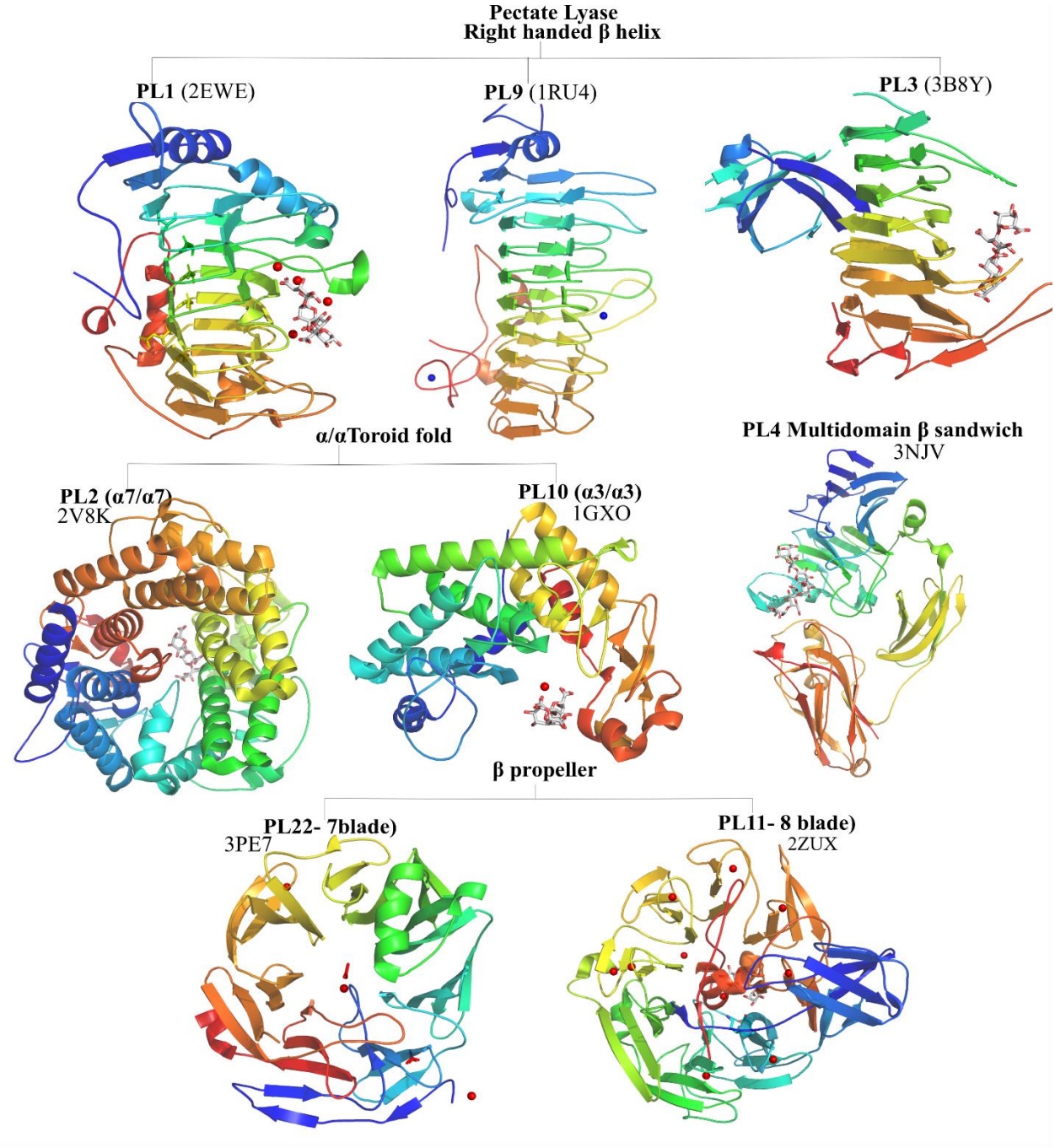

Figure 2:9 Structural Diversity in pectate lyases.

Cartoon diagram of representative structures of pectate lyase with different folds. Their PDB code and the names are shown at the top of each structure. The structures are rainbow colored from blue at the N-terminus to red at the C-terminus. The structure of the enzyme substrate complex is shown for families where complex structures are available. The substrates are shown in stick representation. The top panel shows the right handed β-helix fold seen in PL1, 3 and 9 family. The middle panel shows the (α/α) toroid fold seen in PL2 and PL10 family. The right middle panel shows the multidomain β-sandwich fold seen in PL4. The bottom panel shows the β-propeller fold seen in PL11 and PL22 family.

pectate lyases are virulence factors of plant pathogens. PelA, PelB, PelC, PelD, PelE and PelZ are PL1 virulence factors from *Erwinia chrysanthemi* (Hassan et al., 2013). **PL3** members are endo-lyases with substrate preference for methylated pectin and are the shortest β-helix folds with only eight turns (Akita et al., 2001). Unlike PL1 and PL9 families, the structural rigidity of PL3 members (PelI from *Erwinia chrysanthemi*) is maintained by disulfide bridges and an inner stack of aliphatic interactions (Creze et al., 2008). **PL9** has both exo- and endo-pectate lyases of bacterial and eukaryotic origin. The endo-lyases are mostly extracellular and degrade pectate into large oligogalacturonates and the exo-lyases are in the periplasmic space and degrade oligogalacturonates into unsaturated digalacturonates. An endo-pectate lyase, PelN from *Erwinia chrysanthemi* shows an unusual pH optimum at neutral pH (in contrast to the pH optimum of ~8-10 seen in extracellular pectate lyases) and uses Fe^{2+} as a cofactor (Jenkins et al., 2004; Hassan et al., 2013).

2.4.3.2 $(\alpha/\alpha)_n$ toroid fold in pectate lyase

Another fold observed in the pectate lyase family is the $(\alpha/\alpha)_n$ toroid fold. **PL2 and PL10** belong to this fold. The **PL2** pectate lyases with the $(\alpha/\alpha)_7$ toroid are divided into two subfamilies based on cellular localizations, mode of activity and metal selectivity. Subfamily 1 includes periplasmic, endo-acting lyases with high specificity for homogalacturonans (HGA) (e.g. YePL2A from *Yersinia enterocolitica*). Plasticity in the metal cofactor selection in subfamily 1 reflects the periplasmic localization (multi metal environment) with a pH optimum of 9.6. Subfamily 2 includes cytoplasmic exo-acting lyases with specificity for trigalacturorans and a pH optimum of 8.5. Subfamily 2 have selective metal affinity for Mn^{2+} ion (e.g.YePL2B from *Yersinia enterocolitica*) (Hugouvieux-Cotte-Pattat et al., 2014; Abbott and Boraston, 2008). YePL2A and YePL2B show significant sequence conservation (37% identity) and revealed the presence of an extra loop in the cytoplasmic exo-acting lyase subfamily 2-YePL2B. Mapping of the loop in the structure of YePL2A subfamily 1 places the loop at one subsite of the active site (Abbott and Boraston, 2007). Domain swapping experiments, of the loop region between YePL2A (subfamily 1) and YePL2B (subfamily 2) reduced the rate of HGM degradation by periplasmic subfamily 1 enzyme. This is evidence that the loop contains residues which form the subsite for longer substrates. VvPL2 from *Vibrio vulnificus* YJ016, is an endo-exo transition lyase classified in subfamily 2 (McLean et al., 2015). VvPL2 can degrade both HGM and trigalacturonans with more

specificity for the latter. The presence of a tryptophan residue at the entrance to the active site is responsible for the exo- mode of action. In YePL2A (subfamily 1), endo-acting lyases, the tryptophan at the active site gate is replaced by a lysine. Mutation of the Trp to Lys and Ala in subfamily 2-YePL2B changes its mode of action from exo- to endo-. This is evidence that the gating mechanism responsible for exo-/endo-activity is a single amino acid (McLean et al., 2015).

The pectate lyases of the family **PL10** are found in soil, plant and gut bacteria. PL10 enzymes from *Cellvibrio japonicus* adopt a $(\alpha/\alpha)_3$ barrel topology. The PL10 family enzymes are endo-lytic with substrate specificity for HGA. They show the usual properties of pectate lyases that is an alkaline pH optimum and a Ca^{2+} dependence for enzyme activity (De Armas et al., 2004; Charnock et al., 2002; Hugouvieux-Cotte-Pattat et al., 2014). This is the first family showing convergent evolution of the catalytic site in PL family enzymes (Charnock et al., 2002).

2.4.3.3 β-propeller fold

PL22 family **(Oligogalacturonate lyases (OGLs))**. OGL adopts the seven bladed β-propeller fold and a tunnel like active site. Each blade has four antiparallel β sheets. OGL shows different preferences for localization, substrate specificity and metal cofactor compared to PL1, 3, 9 pectate lyase. OGL are cytoplasmic lyases with substrate specificity for oligogalacturonates. OGL has a pH optimum of 7.3–7.7, which is consistent with their cytoplasmic localization. They use Mn^{2+} or some other transition metal (Co^{2+}, Ni^{2+} or Fe^{2+}) as a cofactor instead of Ca^{2+}. In OGL, a histidine residue acts as the proton abstractor (Abbott et al., 2010).

PL11 (Rhamnogalacturonan lyase-RGL) – PL11 family members are mostly of bacterial origin and break the glycosidic bond between Rha and GalA in the hairy region of pectate. Structural data are available for endo-acting (YesW) and exo-acting (YesX) lyases from *Bacillus* species. Both YesW and YesX fold into an 8 bladed β-propeller with an N terminal extension which folds into two stacks of three-stranded antiparallel β-sheets. Each blade in the propeller has four antiparallel β-strands and except one blade, all other blades possess Ca^{2+} ion which plays mainly a structural role. The active site is present within the deep cleft in the center of the propeller. Structural studies suggest that an extra loop over the active site in the YesX is responsible for its exo-mode of action (Ochiai et al., 2007; Ochiai et al., 2009).

2.4.3.4 Multidomain β-sandwich fold

PL4 (Rhamnogalacturonan lyases-RGL) – PL4 has members of bacterial, fungal and plant origin. RGL from *Aspergillus aculeatus* is a multidomain protein with a new modular architecture and is composed predominantly of β-strands. The N-terminal domain is the largest domain with two antiparallel β-sheets and possesses the catalytic residues. The central and C-terminal domain are involved in substrate binding. RGL is an endo-lyase and degrades RG-I having a minimum length of 12 saccharides. Conserved arginine residues on the multi-domain surface are evidence of an elongated substrate binding site, which can accommodate 12 residues. Ca^{2+} is observed in the structure and mainly plays a structural role. RGL has a pH optimum of 6 and cleaves the glycosidic bond between Rha and GalA (McDonough et al., 2004; Jensen et al., 2010).

2.5 Catalytic Mechanism

As elaborated in the previous section, six structural folds (β-helix, β-jelly roll, $(\alpha/\alpha)_n$ toroid, multidomain β-sandwich, β-propeller and triple-stranded β-helix) have been observed among the 27 PL families. In spite of the diverse substrate specificity and tertiary folds, the degradation of uronate substrate through β- elimination (Figure 2.2 and section 2.1.2) is conserved across the 27 PL families. The observed β- elimination mechanism is either His/Tyr mediated catalysis or a metal assisted elimination mechanism.

Briefly, in the β-elimination mechanism, the negative charge of the C6 carboxylic group (of the uronate substrate) is neutralized either by a metal ion (mostly Ca^{2+}) or by an amino acid (mostly Asn, Asp, Gln or Glu side chains). This neutralization reduces the pKa of the C5 proton and makes it easy for abstraction by a base. Finally, an acid donates a proton to the leaving group. This results in cleavage of the scissile bond between O1-C4 and the generation of an unsaturated product in the newly formed non-reducing end of the sugar substrate.

The lytic mechanism can be classified into two types: (1) the *syn* elimination mechanism, where the abstracted proton and the C4 bridging oxygen are on the same side of the uronate ring and (2) the *anti* elimination mechanism, where the abstracted proton and C4 bridging oxygen are on opposite sides of the uronate ring (M. L. Garron and Cygler, 2014).The following section will describe the β-elimination catalytic mechanism seen in the PL family enzymes.

2.5.1 Metal mediated elimination mechanism

In the metal assisted elimination mechanism, the negative charge of the C6 carboxylic group is neutralized mostly by **Ca^{2+}** ion. This mechanism is commonly seen enzymes with a right handed β-helix fold. Pectate lyase from family **PL1**, **3**, **9**, **10** were shown to utilize this type of elimination mechanism. These are pectate lyases secreted into the extracellular matrix or periplasm. Depending on their intracellular localization, **PL2** members were reported to use either **Mg^{2+}** or **Mn^{2+}** to degrade the highly anionic pectate polysaccharide (Abbott and Boraston, 2007; McLean et al., 2015). Arginine has been observed to act as a catalytic base in PL1, 2, and 10 (Herron et al., 2000; Abbott and Boraston, 2007; Charnock et al., 2002). Here, the enzyme pH optimum is ~10.5 and the Arg interacts with carboxylate group residues in the active site, which causes a shift in the pKa of Arg (~2 units) and makes it a potential base (Charnock et al., 2002). In PL3 and PL9, a Lys residue plays the catalytic base role (Creze et al., 2008;Jenkins et al., 2004). In all these families, a water molecule plays the role of catalytic acid (Garron and Cygler, 2010).

Interestingly, this reaction mechanism has been reported in the **PL6** enzymes ChonB, a GAG lyase and AlyGC an alginate lyase. The amino acids playing the catalytic base and acid roles are Arg and Lys, respectively (Huang et al., 1999; Michel et al., 2004; Xu et al., 2017). **PL22** OGL, an exo-lyase in the cytoplasmic compartment utilizes a Mn^{2+} ion as neutralizer and a His residue as the putative base (Abbott et al., 2010).

2.5.2 His/Tyr mediated elimination mechanism

The His/Tyr mediated elimination mechanism is seen in GAG lyase and alginate lyase (PL5, 7, 14, 15, 17 & 18) family enzymes. These lyases degrade C-5 epimeric substrates, (GlcA/IdoA in GAG and GulA/ManA in alginate) which arises from their adoption of a His/Tyr catalytic apparatus, allowing them to carry out either *syn* or *anti* mediated elimination reactions (Garron and Cygler, 2010). Based on the structural studies on PL family enzymes, this mechanism can be sub-classified into two types: 1) **Tyr/Tyr** mechanism where one Tyr residue plays both catalytic base and acid role. 2) **His (Tyr')/ Tyr** where a His or different Tyr plays the catalytic base role. In both subtypes the neutralization of negative charge on the C-6 carboxylate group is performed by Arg/Asn/Asp/Gln/Glu (M.-L. Garron and Cygler, 2014).

GAG lyase - PL8 family ChonAC and HyL enzymes degrade GlcA containing GAG through *syn* elimination. In ChonAC, Tyr248 is identified to play both the catalytic base and acid role. This dual role of the tyrosine was confirmed by kinetic studies (Rye et al., 2006). ChonABC in PL8 can depolymerize both CS (GlcA) and DS (IdoA). Through a structural comparison with ChonAC and mutational studies, the putative catalytic residues were identified in ChonABC. Tyr461 plays the dual role of proton abstractor and donor for CS-4 or CS-6 (GlcA) substrates through *syn* elimination. To degrade DS (IdoA), however, His345 and Tyr461 perform the base and acid function, respectively. The degradation reaction proceeds in an *anti*-manner (Shaya et al., 2008). HepII from the PL21 family is similar to ChonABC and degrades both GlcA/IdoA containing HS. Like ChonABC, HepII utilizes Tyr257 as catalytic base and acid to degrade the GlcA substrate. For degrading IdoA-containing substrates, His202 and Tyr252 play the role of catalytic base and acid, respectively (Shaya et al., 2010). HepI from PL13 family degrades IdoA containing HS. HepI degrades the substrate by *anti* elimination using His151 and Tyr357 as base and acid (Han et al., 2009). In general, GAG lyase adopts the **Subtype 1 (Tyr/Tyr)** to degrade GlcA containing substrates and the **subtype 2 His (Tyr'/Tyr)** to degrade IdoA containing substrates.

In the **alginate lyase, Subtype 1 (Tyr/Tyr)** is seen in PL5 (A1-III) and PL18 family (aly-SJ02) enzymes. In AI-III, Tyr246 plays both catalytic base and acid role in a *syn* fashion (Yoon et al., 2001), whereas aly-SJ02 uses Tyr353 as catalytic base and acid (Dong et al., 2014). The **subtype 2 His (Tyr'/Tyr)** mechanism is reported in PL7 (A1-II') and PL15 (Atu3025) families. In A1-II', His191 and Tyr284 play the catalytic base and acid role, respectively (Ogura et al., 2008). Atu3025 from PL15 utilizes His311 and Tyr365 as the catalytic base and acid, respectively (Ochiai et al., 2010). Alginate lyase Alg17c from PL17 displays the **(Tyr'/Tyr)** mechanism, where Tyr450 plays the general base abstracting the C-5 proton and Tyr258 plays the role of an acid (Park et al., 2014).

2.5.3 Conformational changes favouring catalysis

Three types of conformational changes have been identified from different crystal structures of PL enzymes. These conformational changes assemble the catalytic residues and facilitates the open and closed states of the active site for catalysis.

1) Lid loop open/close movement – This feature is mostly noticed in the β-jelly roll fold and the single domain $(\alpha/\alpha)_n$ toroid fold. This movement is reported in alginate lyase PL5, PL7, PL18 and GAG lyase PL13 (HepI) (as reviewed in Xu et al., 2018; Han et al., 2009).

2) Domain rotation movement - Upon substrate binding, the multidomain protein undergoes rigid body movement resulting in a rearrangement of residues around the active centre to interact with the ligand. This movement is seen in PL8 family GAG lyases (HyL) (Rigden et al., 2006). This is also seen in alginate lyase PL15 and PL17 families. In particular, the C-terminal β-sheet domain of substrate-bound PL15 alginate lyase undergoes a 10^o rigid body rotation relative to the apo enzyme. Similarly, the β-sheet domain in PL17 alginate lyase undergoes a rigid body rotation of 12^o relative to the apo structure (as reviewed in Xu et al., 2018).

3) Subunit rotation movement - This is seen in PL6 family AlyGC alginate lyase. This enzyme exists as a homodimer and contains an identical active site in each subunit. The complex structure of mutant AlyGC with the ManA tetrasaccharide revealed that the substrate binds the active centre of molecule A, but not that of molecule B. Intriguingly, the initial enzyme-substrate interaction causes a small rotation of molecule B, which enlarges the active site in molecule A with concomitant narrowing of the active site in molecule B. This widening of the (molecule A) active site facilitates substrate binding (Xu et al., 2017).

2.6 New Polysaccharide lyase family - Ulvan Lyase

Ulvan, the cell wall polysaccharide of green algae represents a huge potential biomass. Unlike other marine polysaccharides, alginate (from brown algae), agar and carrageenans (from red algae), ulvan polysaccharides have been less explored (Colle et al., 2011). The development of applications for ulvan polysaccharide will require specialized enzymes to process this polymer. Gram-negative bacteria isolated from mud were the first to show the presence of ulvan degrading activity. Based on the reaction product being an unsaturated uronic acid (4-deoxy-L-*threo*-hex-4-enopyranosiduronic acid, ΔUA), the corresponding enzyme was identified as a lyase (Lahaye et al., 1997). The lyase enzyme from a filamentous fungus *Trichoderma* sp. GL2 was also reported to utilize ulvan as a substrate, but its main substrate was identified to be oligoglucuronan (Delattre et al., 2006). Only recently, two ulvan degrading enzymes were identified in the Gram-negative bacterium *Nonlabens (Persicivirga) ulvanivorans* strain PLR (NLR42, NLR48) (Colle et al., 2011). The identified genes were cloned, purified and biochemically characterized. They both

degrade ulvan in an endo-lytic fashion. The protein sequence of this novel lyase was unrelated to any other lyases and became the first representative of a new lyase family called PL28 (Colle et al., 2011). Subsequent genome sequencing of marine bacteria identified many novel ulvan lytic enzymes. The genome of *N. ulvanivorans* was recently sequenced (Kopel et al., 2014a), followed by sequencing the genomes of *Pseudoalteromonas* sp. Strain PLSV (Kopel et al., 2014b) and *Alteromonas* sp. strains LTR and LOR (Kopel et al., 2014c). This led to the identification of several putative ulvan lyase genes. The identified putative ulvan lyase gene clustered with genes predicted for β-glucuronyl hydrolase, sulfatase, rhamnosidase and xylosidase activities. It was proposed that this gene cluster coordinates the complete saccharification of ulvan polysaccharide (such gene clusters are called Polysaccharide utilization Loci (PULs)) (Foran et al., 2017). Ulvan PULs in *Alteromonas* strain LOR, *Pseudoalteromonas* strain PLSV and *Nonlabens* are shown in **Figure 2.10** and the putative ulvan lyases are highlighted in a purple box. These newly identified ulvan lyases are classified under PL24, PL25 and PL28 family in the CAZy database (Lombard et al., 2014). In addition to ulvan lyases, the first ulvan hydrolase from *N. ulvanivorans* was characterized. The enzyme showed unsaturated β-glucuronyl activity and works together with ulvan lyases to fully depolymerize the ulvan polymer (Collén et al., 2014).

Many of these putative ulvan lyase genes were recombinantly expressed and partially characterized. LOR107 from *Alteromonas* classified in the PL24 family was reported to act endolytically on ulvan polysaccharide. The end product of the degradation reaction is predominantly a tetrasaccharide with IdoA (Rha3S-IdoA-Rha3S-ΔUA) (Kopel et al., 2016). This shows the preference of LOR107 for cleavage next to GlcA in the ulvan. Currently, there are 30 and 27 sequences of bacterial origin classified in the PL24 and PL25 families respectively.

Biochemical characterization of the first identified ulvan lyase NLR42 (NCBI accession number: AEN28574, 46 KDa with two domains) and NLR48 (NCBI accession number: KEZ94336) from *Nonlabens ulvanivorans* strain PLR (PULs is not identified) showed that these enzymes endolytically degrade ulvan. The end product is a mixture of tetrasaccharides comprised of IdoA and GlcA (Rha3S-GlcA-Rha3S-ΔUA and Rha3S-IdoA-Rha3S-ΔUA). This shows that NLR42 and the homolog of its catalytic module, can cleave ulvan next to both GlcA and IdoA. The PL28 family has presently 18 sequences of bacterial origin.

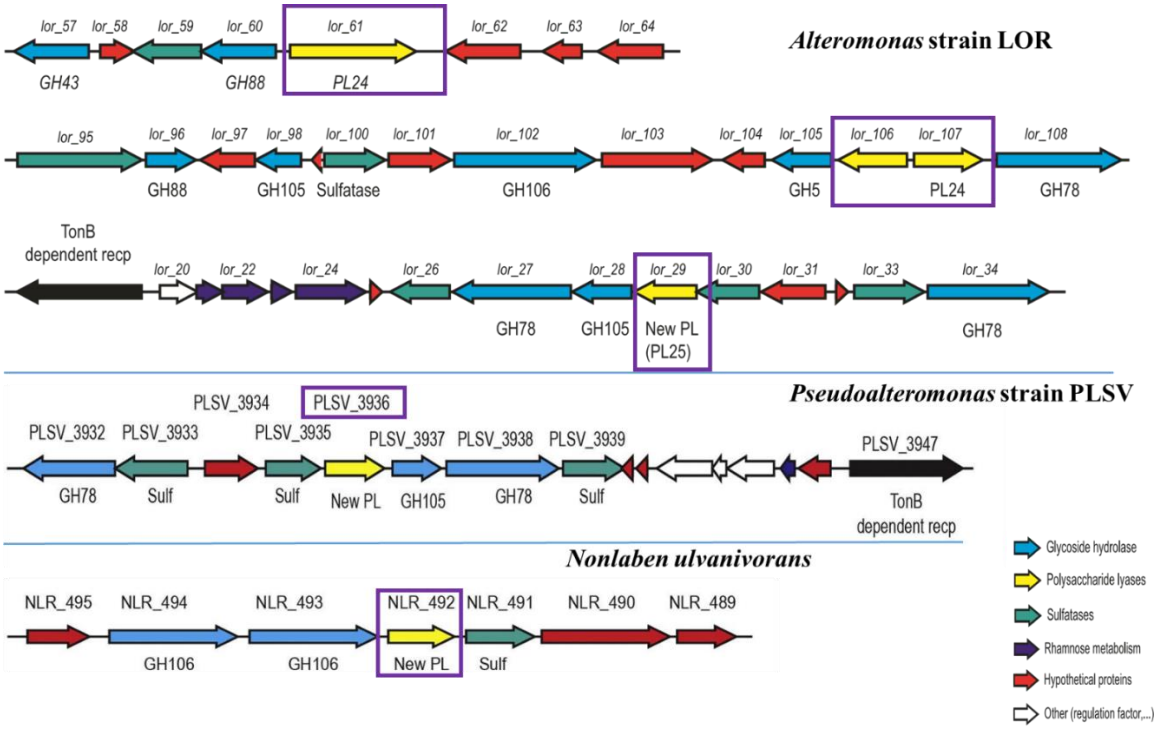

Figure 2:10 Ulvan PULs identified in marine bacteria.
PULs identified in *Alteromonas* strain LOR, *Pseudoalteromonas* strain PLSV and *Nonlaben ulvanivorans*. Hydrolases (colored blue), lyases (colored yellow), sulfatases (colored green) and their color code is shown in the insert. The enzymes highlighted in purple boxes are the lyases, we have received from collaborators. Figure adopted and modified from (Foran et al., 2017)

Ulvan lyases have not been structurally characterized and their catalytic mechanism is unknown. It is therefore of interest to determine if the ulvan-degrading enzymes display a novel architecture and what catalytic mechanism and apparatus they utilize.

3. Thesis rationale and objective

Studying the structure and catalytic mechanism of bacterial GAG lyases has been a long standing interest in our laboratory. The structure of several chondroitin lyases, chondroitinase B, heparinase I (HepI) and heparinase II (HepII) have been previously solved in our laboratory. These structural studies revealed several interesting features of the fold and allowed us to propose the catalytic mechanism.

HepI degrades the highly sulfated heparin (HP), whereas HepII degrades both HP and less sulfated heparan sulfate (HS). As mentioned in the previous section, HepIII, a member of the PL12 family, degrades the heparan sulfate (HS) substrate. All three heparinase enzymes share limited sequence similarity. When the candidate joined the laboratory, the structures of HepI and HepII had already been determined and published but the structural studies of HepIII were in progress. Although the structure of the apo HepIII had already been determined in our lab, additional investigations of HepIII were required to detail its catalytic mechanism. The **objective** of this project is to obtain the HepIII/substrate complex structure to understand the catalytic mechanism. We **hypothesized** that **HepIII** might utilizes the His/Tyr catalytic mechanism as seen in other GAG lyases. The major efforts in this research were directed towards: **1)** identification of catalytic residues, **2)** generation of putative active site mutants and measuring their enzyme activity, **3)** crystallization of wild type and active site mutants of HepIII with the tetrasaccharide substrate.

At about that time, a new family of PL enzymes with the ability to degrade a marine polysaccharide called **ulvan** was discovered. Several potential ulvan lyases were identified in marine bacteria through large-scale genome sequencing efforts of marine organisms. These lyases were classified into three new sequence families in the CAZy database and there was no structural information on any of them. In collaboration with the groups that identified and characterized the first ulvan lyases we engaged in their structural studies. The **objective** of this project is to expand the structural and functional information on polysaccharide lyases by providing structural information on three new PL lyase families: PL24, PL25 and PL28. We **hypothesized** that the ulvan lyases might evolved to have a different scaffold (fold) but preserves the β-elimination catalytic machinery. The efforts of the candidate were focused on: **1)** determination of the atomic structure of ulvan lyases: LOR107, PLSV3936 and NLR48, **2)** identification of putative active site residues by the generation and testing of mutants, **3)** obtaining the structures of enzyme/substrate

complexes to determine the substrate binding mode, recognition of the specific oligosaccharide sequence and the catalytic mechanism.

Table 3-1 Knowledge gap in the structural studies of PL family enzymes

PL	Family	Organism	Substrate	Fold	Mechanism
12	Heparin lyase III	*Bacteroides thetaiotaomicron*	Heparan sulfate	$(\alpha/\alpha)_6$ toroid + anti-parallel β-sheet	*
24	LOR107	*Alteromonas* sp. LOR	Ulvan	*	*
25	PLSV3936	*Pseudoalteromonas* sp. PLSV	Ulvan	*	*
28	NLR48	*Nonlabens ulvanivorans*	Ulvan	*	*

* - Uncharacterized

4. Methods and materials

This chapter is divided into three sections. The first section will describe the steps in crystallography and briefly explain the methods in each step. The second section of this chapter will describe the experimental procedures followed in specific projects. The third section will outline the materials (list of primers, plasmids, expression cell lines and list of all chemicals) used in this research.

4.1 Steps in protein crystallography

4.1.1 Cloning

The first step in determining the three-dimensional structure of a protein by crystallography is to produce a sufficient quantity of highly purified target protein. For protein crystallization one needs a few milligrams of protein sample at >95% purity. Typical problems at this step are low expression of the target protein, poor solubility and difficulty in removing contaminating proteins. In addition, the protein may form soluble aggregates, precipitate or degrade during or after purification. Additional difficulties may occur for eukaryotic proteins as they may require post-translational modifications for proper folding. The advancement in recombinant DNA technology, coupled with the design of more efficient and diverse expression systems, have made these problems tractable. An adjustment of a growth parameter such as time, growth temperature and inducer concentration is frequently sufficient to decrease the fraction of insoluble protein and to increase the yield. Other strategies to produce soluble, well-behaving protein are: 1) to fuse the protein with tags like MBP (Maltose Binding Protein) or GST (Glutathione S-transferase), which have been demonstrated to help in protein folding (Bell et al., 2013), 2) to make truncations of the protein based on secondary structure predictions by removing the unstructured regions, 3) to change the expression system, e.g. for expression in *E. coli* there are several specialized *E. coli* strains that might improve the yield and quality of expressed protein such as *E.coli* strains carrying chaperones to assist in protein folding (Makhoba and Mthembu, 2015) or strains with additional tRNAs for rare codons (Rosano and Ceccarelli, 2014).

4.1.2 Quality assessment of protein sample for crystallization

Quality of the protein in terms of both **purity** and **homogeneity** determines its crystallization propensity. The protein sample should be homogeneous (meaning a single oligomeric state) to increase the chances of crystallization. Sometimes the purified protein tends to form soluble aggregates or precipitates and degrades. Many parameters like temperature, ionic strength and the pH of the buffer influence homogeneity of the sample. Upon obtaining soluble protein, the purity and homogeneity of the sample is evaluated by SDS-PAGE, size exclusion chromatography (SEC) and dynamic light scattering (DLS). Troubleshooting a heterogeneous sample is often carried out by changing the buffer, pH, ionic strength or by adding chemical additives. The effects of these changes are monitored by DLS. In this case, the DLS experiment is

very helpful because it consumes a limited amount of protein and allows the simultaneous screening of multiple conditions (96 well format) in a short time (Proteau et al., 2010). Protein degradation can be eliminated by the addition of protease inhibitors in all the steps during purification or by engineering the site susceptible to proteolysis. (Lebendiker and Danieli, 2014; Wingfield, 2015)

4.1.3 Crystallization

The crystallization phenomenon involves changing the solubility of protein to attain the supersaturated state. Under favorable conditions, protein molecules form small, ordered nuclei that assemble into crystals. Many parameters such as precipitant, buffer, pH, ionic strength, protein concentration and temperature affect protein solubility. Thus, these parameters are varied in the crystallization experiment to guarantee that the solubility limit is reached.

The crystallization **phase diagram** helps to predict the change in solubility as a function of different experimental parameters. The phase diagram consists of four zones: the **high supersaturation** zone where protein precipitates; the **moderate supersaturation zone**, which favors nucleation events; the **metastable zone** where the protein crystals may grow big and the **undersaturation** zone where the protein concentration is below saturation and there is no chance for its crystallization (Rupp, 2015; Chayen, 2004) **(Figure 4.1)**.

Challenges in protein crystallization are offset to some extent by tremendous improvements in recombinant DNA technology, protein production, purification and quality assessment. The process and parameters influencing protein crystallization are also well understood. But the condition under which a specific protein will crystallize is not yet predictable. It still remains a trial-and-error screening approach, using a wide array of chemical conditions (Garman, 2014).

Some proteins are recalcitrant to crystallization and this may be an inherent property of the protein. In such cases, rescue strategies like surface entropy reduction, lysine methylation, limited proteolysis or cloning different constructs by removing the predicted unstructured flexible regions are potentially useful. Chemical modification of surface lysine residues reduce the protein solubility and modify its surface properties, which may promote crystallization. (Walter et al., 2006; Sledz et al., 2010).

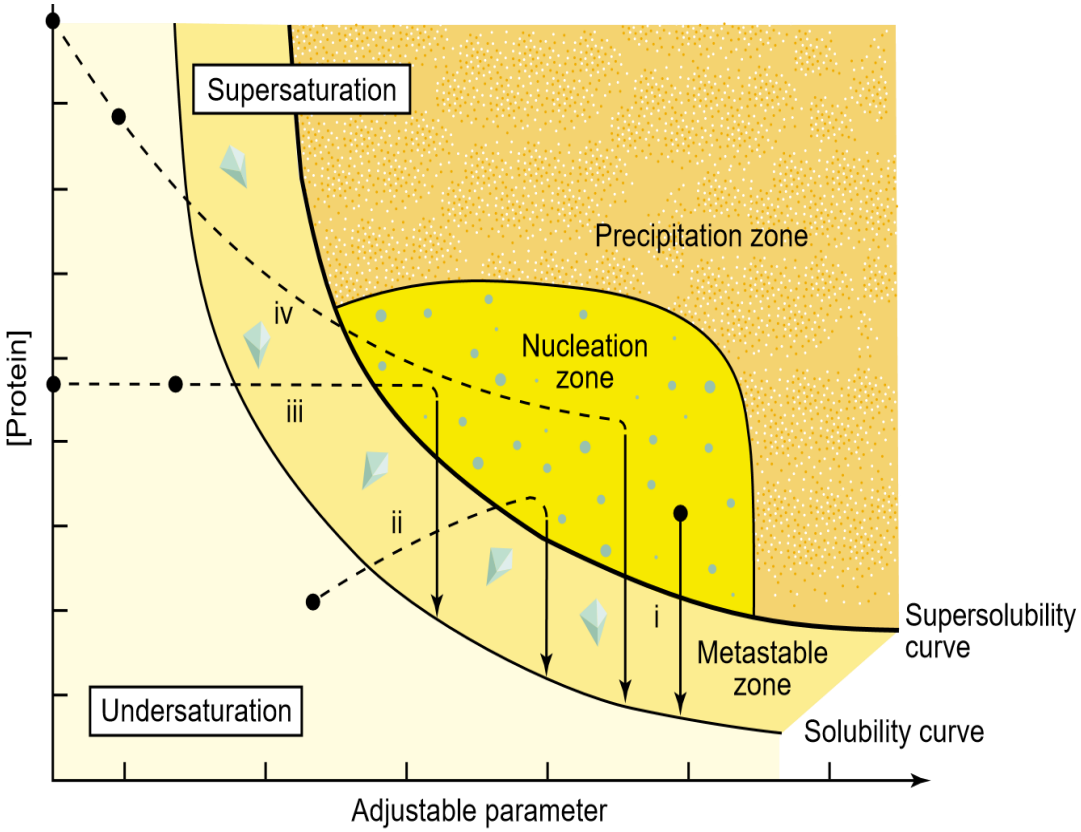

Figure 4.1:1 Phase diagram for protein crystallization.
The supersaturation zone where protein precipitates; the nucleation zone where the nucleation event occurs; the metastable zone where crystals are grown due to lower supersaturation; the under-saturation zone where protein crystals does not grow (Figure adapted from (Chayen, 2004)).

4.1.3.1 Crystal growth optimization

The initial crystallization conditions obtained from broad screening must often be optimized to produce bigger and better diffracting crystals. In this step, conditions are varied in a finer grid by varying protein concentration, buffer, pH, ionic strength and precipitant concentration around the initial conditions. Additives (divalent metal ions and other chemicals) are often used in this step to improve the crystal quality. **Seeding** is a powerful manipulation technique in the optimization step that disconnects nucleation from crystal growth. In microseeding, the crystals are crushed and added to a fresh crystallization drop. These invisible crystals act as nucleation sites for crystal growth. In macroseeding, a single crystal is transferred to the pre-equilibrated drop (in the metastable zone). This crystal serves as a single point of nucleation, whereupon protein molecules from the drop can arrange themselves on the existing crystal and make it grow bigger. (Till et al., 2013; Shaw Stewart et al., 2011).

4.1.3.2 Crystallization of protein-ligand complex

The structure of a protein-ligand gives more information (e.g. active site, residues crucial for catalysis, residues facilitating substrate binding) than the apo protein structure alone. The protein-ligand complex can be captured by using a non-hydrolysable substrate analog or by using an inactive catalytic site mutant (here the reaction cannot proceed and the substrate will stay in the active site). The ligands can be introduced either by co-crystallization or soaking. In co-crystallization, the protein and ligand are mixed prior to crystallization. This method usually requires searching for new crystallization conditions. The other option is to introduce the ligand into a preformed protein crystal by immersing the crystal in a ligand-containing solution and relying on diffusion of the ligand into the crystal. Protein crystals are composed of ~50% to 70% water and the ligand can diffuse into the protein molecules in the crystal through solvent channels. A disadvantage of soaking is that the introduction of ligand can disintegrate the protein crystal because of the change in chemical environment. Soaking might also disturb the crystal lattice and packing, ultimately affecting the diffraction negatively. The concentration of ligand to use in the co-crystallization and soaking experiment and the time of ligand soaking (diffusion time) are the major parameters to work through (Hassell et al., 2006; Müller, 2017).

4.1.4 Cryo-crystallography and data collection

Preparation of the crystal for exposure to a very intense X-Ray beam is a crucial step in the experiment; the protein crystals are very delicate and are easily damaged by radiation. The quality of diffraction data determines the level of structural details we obtain from the structure. Before irradiating with X-rays, the cryo-protected crystals can be flash cooled in liquid nitrogen and the experiment performed at low temperature, usually ~100K, to minimize radiation damage (Rodgers, 1994; Pflugrath, 2015). The critical step here is choosing the appropriate cryogenic conditions to prevent ice formation within the crystal. Also, other experimental parameters like X-ray exposure time, crystal to detector distance and oscillation degree need to be taken into account. Frozen crystals are mounted on the goniostat and placed in the X-ray beam. The crystal is kept at low temperature during the experiment by being immersed in a N_2 gas stream cooled to 100K. The crystal is rotated during data acquisition and the diffraction intensities in various spots of the flat detector surface are integrated over a small rotation angle and stored in separate files (frames). The quantities to be followed during the data collection experiment are the number of measured

reflections out of all possible reflections (completeness) and how many times each reflection was recorded (redundancy) (Wlodawer et al., 2013; Krojer et al., 2013).

4.1.5 Data processing

The computational part of the crystallography experiment starts with data processing. The result of diffraction experiment is a large dataset containing information about the intensities and positions of the scattered X-rays (reciprocal space). In data processing, the measured intensities are converted into structure factor amplitudes. This involves many intermediate steps. **Indexing** - involves identification of spots in the diffraction frame and assigning Miller indices (hkl) to each reflection, followed by refinement of the crystal and instrument parameters to best fit the observed positions of the spots to predicted ones. This step provides the unit cell parameters and the Bravais lattice of the crystal. **Integration** - The intensity of each reflection is obtained by summation of intensities over all pixels considered to be part of the diffraction spot and the background noise is subtracted. These raw intensities are corrected for distortions due to the experimental setup. The next step is **scaling** and **merging** of redundant, symmetry related reflections into a unique dataset. Several algorithms and software packages are available for data processing. (Wlodawer et al., 2008; Powell, 2017)

The quality of the diffraction data is assessed through the following parameters: data redundancy, merging R-values, data completeness, $\langle I/\sigma(I) \rangle$ (signal-to-noise ratio) and CC1/2 (Wlodawer et al., 2013; Wlodawer et al., 2008). In the diffraction experiment, each reflection is measured multiple times because of the symmetry in the crystal and the oscillation method of data collection. **Redundancy** is the number of times one unique reflection is measured. Each reflection intensity is measured with some error and the data accuracy can therefore be improved by increasing redundancy. **R-merge** reflects the difference between symmetry related reflections and is calculated using equation 4.1. For a good dataset, a low discrepancy between symmetry related reflections coincides with a small value of $R_{merge.}$ For a suboptimal dataset the R_{merge} would be >10% and this value tends to rise slightly with higher redundancy. R_{merge} can be artificially manipulated by omitting weak or partial reflections and less redundant data, but this will eliminate the useful high resolution reflections. Hence the modified R parameters $\mathbf{R_{meas}}$ (redundancy independent merging R factor) and $\mathbf{R_{pim}}$ (precision indicating merging R factor) are taken into consideration while assessing data quality (Wlodawer et al., 2008). Data **completeness** above 95

% and 75 % should be measured in the overall and highest resolution shell respectively (Dauter, 2017). The **resolution** cutoff for the dataset is primarily decided by the signal to noise ratio ***<I/σ(I)>*** and a value in the range of 1-2 is generally accepted (Evans and Murshudov, 2013). Another statistical parameter to determine resolution cutoff is **CC1/2** and a value above 60% is generally accepted (Karplus and Diederichs, 2012).

$$R_{merge} = \sum_{h} \sum_{i=1}^{N} |I_{(h)i} - \bar{I}_{(h)}| \,/\, \sum_{h} \sum_{i=1}^{N} I_{(h)i} \tag{4.1}$$

Where $I_{(h)i}$ is the ith intensity measurement of a reflection h and $\bar{I}_{(h)}$ is its average intensity.

4.1.6 Phasing

The electron density map (***ρ(xyz)***) of a protein molecule in real space is calculated from the reciprocal space amplitudes by a Fourier transform. This calculation requires two components for each reflection: the structure factor amplitude (**$|F_{hkl}|$**) and its phase (**α_{hkl}**) (shown in equation 4.2). In the diffraction experiment, only the intensities of X-rays scattered from the crystal are measured and the structure factor amplitude can be calculated from the measured intensity using [$|F_{hkl}| \propto \sqrt{(\text{Intensity}_{hkl})}$]. Unfortunately, the phase information is lost in the diffraction experiment, the so-called phase problem in crystallography (Taylor, 2003).

$$\rho(xyz) = \frac{1}{V} \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)] \tag{4.2}$$

One of the major breakthroughs in protein crystallography is the development of methods and techniques to solve the phase problem. The available methods for protein crystallography are: Multiple Isomorphous Replacement (**MIR**) where phasing is done by a heavy atom soak into the native protein crystal, Molecular Replacement (**MR**) where the phase information is extracted from homologs, Multiple-wavelength and Single-wavelength Anomalous Dispersion (**MAD** & **SAD**) method where phases are obtained by the use of anomalous scatterers. Advancement in the methods and algorithms such as density modification (solvent flattening, non-crystallographic symmetry (NCS) averaging) makes **SAD** a routine method to determine new structures of proteins and **MR** to determine structure of proteins where the structure of a homolog (sequence identity ≥ 25%) is known (Taylor, 2010).

4.1.6.1 Isomorphous replacement

The first protein structure, myoglobin was solved in the 1950's by the MIR method using heavy atom soaks (Kendrew et al., 1958). For this method, it is critical that the soaking process does not significantly affect the unit cell parameters or the orientation of molecules, such that isomorphism between the native and derivatized dataset is maintained. The heavy atom causes measurable change in intensities during the diffraction experiment. These intensity changes are used to find the heavy atom positions. Then the initial phase estimation of the native protein is calculated by comparing the intensities of (non-derivative) native and (derivatized) heavy atom datasets. This method gives two solutions. To overcome this apparent phase ambiguity, usually more than one derivatized dataset is needed. With the help of solvent flattening and density averaging methods, however, it is possible to obtain initial phase estimates from a single derivative dataset (Taylor, 2010; Taylor, 2003).

4.1.6.2 Molecular replacement (MR)

Initial phase estimates are obtained using the structure of a homologous protein, which has ≥25% sequence identity with the unknown structure, as a model. The proper orientation of this within the unit cell of the unknown protein is determined through a search for best correlation of the Patterson functions of the model and unknown protein as a function of the orientation of the model. In the second step the properly oriented model is translated across the unit cell of the unknown protein to find the position of the model that gives the highest correlation between the calculated and observed structure factors. Combining the phases from the model with the observed structure amplitudes allows us to calculate an electron density map that has somewhat distorted features of the unknown molecule and provides a good starting point for refinement (Evans and McCoy, 2008).

4.1.6.3 Single anomalous dispersion (SAD)

When an atom is irradiated at a wavelength far from its absorption edge, the scattering is predominantly elastic. However, when the X-ray wavelength is at the absorption edge of an atom, the inelastic scattering becomes measureable. This is related to the excitation of an electron from a low energy orbital to a higher energy shell followed by its return to the ground state. This phenomena results in anomalous scattering and causes a 90^{o} phase change of the emitted photon.

This difference in the diffraction pattern is used to estimate the location of the anomalous scatterers and, in a manner similar to SIR (Single Isomorphous Replacement), to derive estimates of the phases of protein structure amplitudes. The most commonly used anomalous scatterer is selenium. The popularity of this relatively light atom is due to an easy incorporation of selenium into the protein as a selenomethionine (SeMet) in place of methionine during recombinant protein expression. In the SAD experiment, the SeMet incorporated crystals are irradiated at a wavelength corresponding to the selenium absorption edge (Taylor, 2010).

4.1.6.4 Selenomethionine (SeMet) labelling of recombinant proteins

Two alternative methods are frequently used to incorporate SeMet into the protein of interest: 1) using an *E. coli* methionine auxotroph mutant as an expression strain (Hendrickson et al., 1990) or 2) by inhibiting the methionine biosynthesis pathway during protein expression (Walden, 2010). In the first method, the auxotroph cells only grow in the medium supplemented with methionine or SeMet. In the second method, the methionine biosynthetic pathway is inhibited by adding hydrophobic amino acids which inhibit aspartokinase, the first enzyme involved in methionine biosynthesis. By supplementing with SeMet, these cells, which lack the ability to produce their own methionine, are forced to incorporate this alternative amino acid into their newly synthesized proteins.

4.1.7 Model Building and Refinement

The final result from a protein crystallography experiment is the **electron density map.** By applying the inverse Fourier transform to structure factors (amplitude and phase) and through density modification (solvent flattening and symmetry related density averaging) the initial electron density map is obtained. An initial protein model can be built into the density map using specialized crystallographic software. The initial electron density map and model are not accurate and need to be improved by several rounds of refinement. The aim of **refinement** is to minimize the difference between structure factor amplitudes calculated from experimental data and those calculated from the model. It is an iterative process that cycles between automated correction of refinement parameters (e.g. *x, y, z* positional coordinates and B factors or atomic displacement parameters) by a specialized software and manual improvement of the model. The progress of refinement is monitored by statistical parameters called R-factors (R-work and R-free) given by the equation 4.3.

$$R_{work} = \sum_{hkl} [|F_{obs}(hkl)| - |F_{calc}\,(hkl)|] / \sum_{hkl} |F_{obs}\,(hkl)| \quad (4.3)$$

Where $|F_{obs}(hkl)|$ is the observed structure factor amplitude from the diffraction experiment and $|F_{calc}(hkl)|$ is the calculated structure factor amplitude from the model.

To cross validate the refinement process and to avoid overfitting the model, a small percentage of reflections (~5%) are not included in the refinement. The parameter called **R-free** is calculated for these reflections using the above equation (Brünger, 1992). In an ideal refinement process, the R-work and R-free values should attain a minimum value and their difference should be less than ~5% (Wlodawer et al., 2013; Wlodawer et al., 2008).

4.1.8 Electron density map and its interpretation

There are two types of electron density maps commonly used to evaluate the model: the $2F_o$-F_c and F_o-F_c maps. The $2F_o$-F_c map is usually contoured at 1 σ. The F_o-F_c map gives the positive and negative peaks which indicate insufficient or excessive regions of electron density predicted by the model (Wlodawer et al., 2013; Wlodawer et al., 2008). After each round of refinement, both density maps are inspected. Water molecules are added to the structure automatically during the refinement step or manually during rebuilding.

4.1.9 Model validation and submission

The chemical environment of the protein structure is carefully examined during the refinement step. In addition, the refined model is investigated for deviations from the expected chemical, stereochemical and biochemical parameters (e.g. bond length, bond angle, torsion angles, Ramachandran outliers). Validation servers (e.g. procheck, molprobity) are available for this task (Chen et al., 2010). The completed, validated model is then deposited to the Protein Data Bank **(PDB)** where it is made available to other researchers.

4.2 Experimental procedure

4.2.1 Molecular biology and microbiology techniques

4.2.1.1 Ligation independent cloning (LIC)

The protein constructs used in this research were made by following the ligation independent cloning protocol. Briefly, **LIC cloning** uses the 3'→5' exonuclease activity of T4 DNA polymerase to make sticky complementary ends (of ~15 nucleotides in length) in the vector and insert. T4 DNA polymerase has 3'→5' exonuclease activity and 5'→3' polymerase activity. In the absence of nucleotides in the reaction mixture, the exonuclease activity takes over and clears the strand. Interestingly, if we supply more of a single nucleotide (e.g. dGTP), upon encountering nucleotide "C" in the opposite strand, the polymerase and exonuclease activity competes and the enzyme is stalled at this site (Eschenfeldt et al., 2009).

4.2.1.2 PCR components

The vector and insert were amplified by PCR with specific sets of primers using KOD polymerase. The 50 µl PCR reaction mixture consists of ~50 ng template DNA, 1X reaction buffer, 1.5mM $MgSO_4$, 200 µM dNTP mix, 0.3 µM primer pair, 1 unit of KOD hot start polymerase and nuclease free water. The thermocycle protocol of PCR is shown below. This is the standard protocol followed in this thesis. Annealing and extension temperatures were varied based on the primer and the length of DNA to be amplified.

Thermocycle protocol for the PCR reaction

1) Polymerase activation — 95°C for 2 minutes
2) Initial Denaturation — 95°C for 20 seconds
3) Annealing — Average Tm^{o}C-5^{o}C of primer pair for 20 seconds
4) Extension — 70°C for 20 seconds/KB
5) Final extension — 70°C for 5 minutes

Steps 2 to 4 were repeated in ~25 to 30 cycles. In site directed mutagenesis PCR, steps 2 to 4 were repeated for 20 cycles. The amplified vector and insert were analysed and separated in a 1% agarose gel. The DNA of correct molecular weight was excised from the agarose gel and extracted using QIAquick Gel Extraction Kit by following the manufacturer's protocol.

4.2.1.3 T4 Polymerase treatment and ligation

The purified vector and insert DNA were subjected separately to T4 DNA polymerase treatment. The 40 µl reaction mixture consists of ~200 ng DNA, 1X NEBuffer2, 1×NEB BSA, 5 mM DTT, 0.3 µl T4 DNA polymerase and 2.5 mM of a single nucleotide. The nucleotide used in the treatment of vector and insert were dGTP and dCTP, respectively. The reaction was carried out for an hour at 22ºC followed by deactivation of the enzyme at 75°C for 20 minutes. T4 DNA polymerase treated vector and insert were mixed in the ratio of ~ 1:5 to 1:10 and incubated on ice for 30 minutes to allow ligation.

4.2.1.4 Transformation

The ligated mixture was transformed into chemically competent *E.coli* DH5α cells. For transformation, 5 µl of ligated product was incubated with 50 µl of chemically competent *E.coli* DH5α cells on ice for 30 minutes. This was followed by heat shock at 42° C for 30 seconds and rapid cooling on ice for 2 minutes. Five hundred µl of Luria-Bertani broth (LB, 1% (w/v) tryptone, 0.5% (w/v) yeast extract and 1% (w/v) NaCl) was added and the recovery was carried out for an hour at 37ºC. After recovery, the cells were pelleted at 3000 rpm for ~3 minutes and the pellet was plated on an LB agar (supplemented with appropriate antibiotic) and incubated overnight at 37° C. The next day, a single colony was inoculated in LB media and grown at 37° C for ~16 hours. After 16 hours of growth, the cells were spun down and the plasmid was isolated using QIAprep Spin Miniprep Kit following the manufacturer's protocol. The inserts were further confirmed by sequencing at the National Research Council (Saskatoon, Canada) and at the Eurofins (Huntsville, USA) sequencing facility. This transformation protocol was used for all constructs listed in this thesis.

4.2.1.5 Site directed mutagenesis

Several putative active site mutants were made using the Quickchange site-directed mutagenesis protocol and using KOD polymerase. The primer pairs carrying the desired mutations used in the projects are shown in section 4.3. Briefly, the plasmid containing the gene of interest was amplified with the primer pairs carrying the mutation using KOD polymerase as described in section 4.2.1.2. After PCR, the template plasmid was digested with 1 µl of DpnI enzyme for an hour at 37° C. Five µl of PCR product was then transformed into 50µl of chemically competent *E.coli* DH5α cells, as described in section 4.2.1.4. The clones carrying the desired

mutation were confirmed by sequencing at the National Research Council (Saskatoon, Canada) and Eurofins (Huntsville, USA) facility.

4.2.2 Recombinant protein production

Expression plasmids carrying the gene of interest were transformed into chemically competent *E.coli* expression cells lines as described in section 4.2.1.4. For large scale protein production, a single positive colony was inoculated into 25 mL LB media supplemented with appropriate antibiotics and grown overnight at 37° C. This overnight inoculum was subcultured into 1 litre of Terrific Broth media (TB media, 1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 72 mM K_2HPO_4, 17 mM KH_2PO_4, and 0.4% (v/v) glycerol) supplemented with the appropriate antibiotic. Cells were allowed to grow at 37° C until the absorbance at 600 nm reached 1.2-1.5, then the temperature was reduced to 18° C or 20° C and protein expression was induced using 0.5 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). The IPTG induced cells were allowed to grow at low temperature for ~16 hours. After that, the cells were centrifuged at 6200 x g in a JLA8.1 beckman rotor for 20 minutes. The cell pellets were resuspended in a lysis buffer and the protein was purified from them. Sometimes, the cell pellets were stored at -80 °C until further processing could be carried out. The various expression vectors and cell lines used in the different projects are listed in tables 4.2.1 and 4.2.2, respectively.

Table 4.2-1. List of vectors and their description.

Plasmid	Description	Source
pMSCG7	Ampicillin^{resistant} (Amp^r), expression vector with an N-terminal 6xHis tag and a TEV cleavage site	Cygler's Lab
pRL652	Amp^r, expression vector with an N-terminal GST tag and a TEV cleavage site (modified version of pGEX vector adopted for LIC cloning)	Made by Dr. Andrey Grishin at Cygler's Lab
pET28a	Kanamycin^{resistant} (Kan^r), expression vector with a C-terminal 6xHis tag and a thrombin cleavage site	Novagen
pMCSG9	Amp^r, expression vector with an N-terminal MBP tag and a TEV cleavage site.	Cygler's Lab

Table 4.2-2. List of bacterial strains used in the various projects.

Cell Lines	Project
E. coli DH5α	All
E. coli BL21	HepIII
E. coli BL21(DE3)	NLR48
E. coli BL21(DE3)-RIPL	LOR107, PLSV3936

4.2.2.1 Purification of recombinant protein

The resuspended cells were lysed using either a cell disruptor at 35 psi (Constant Systems Ltd, UK) or sonicator (4 minutes, 7 sec on, 15 sec off). The soluble fraction was obtained by centrifuging at 39,000 x g for 40 min using a JA25.50 beckman rotor. Proteins were purified from the soluble fraction by affinity chromatography.

For His tagged protein, Ni-NTA or Cobalt talon resin was used in the purification. The resin was first equilibrated with lysis buffer and the soluble fraction was incubated with resin for an hour. This was followed by a wash step with buffer containing a low concentration of Imidazole to remove contaminants. The target protein was then eluted with a buffer containing higher imidazole concentration.

For GST fusion proteins, glutathione resin was used in the purification. The glutathione resin was pre-equilibrated in lysis buffer and the soluble fraction was incubated with resin for two hours. The unbound contaminants were washed with high salt buffer, followed by on-column GST tag cleavage using His-TEV (Tobacco Etch Virus) protease. The TEV protease was added at a mass ratio of 1:50 and the reaction was carried out for ~8 to 12 hours, either at room temperature or at 4^{o}C. After GST cleavage, the elution fraction containing untagged target protein and His-TEV protease was passed through the Ni-NTA column. This step removed the TEV protease and resulted in pure target protein.

This purification protocol is followed in all the projects described in the thesis with slight modifications based on the protein behavior. The composition of lysis buffer, wash buffer and elution buffer used in the purification of various proteins is mentioned in the following section.

4.2.2.2 Quality assessment of protein sample

All the fractions obtained in the purification steps were analyzed by (SDS-PAGE), which separates the proteins by molecular weight. Protein fractions containing pure samples of correct molecular weight were pooled and examined for oligomeric state by size exclusion chromatography (SEC) using GE S200 or Biorad SEC650 size exclusion columns. In SEC, the samples are separated based on their molecular weight. The sample coming off in the void volume represents large aggregates and is not useful. The target proteins are separated based on oligomer size.

4.2.3 Crystallization trials and crystal growth optimization

After obtaining an adequate amount of homogeneous sample from SEC, the samples were concentrated using an appropriate molecular weight cut-off centrifugal filter at 4000 rpm in 10-20 min intervals. Crystallization experiments were attempted for the concentrated sample. A wide range of commercial screens were used in the search for hit conditions. These screens were of the 96-well, sitting drop format and our screening protocol involved mixing 0.3 µl protein with 0.3 µl reservoir using a Gryphon crystallization robot. The initial crystal hits were optimized by hanging drop vapor diffusion method with seeding. This step is carried out in all projects.

4.2.3.1 Diffraction experiment and structure determination

All the diffraction data described in this research were collected at the CMCF (Canadian macromolecular crystallography facility) beamline (ID or BM), Canadian Light Source (CLS). Details of the data collection and processing are described in the following sections for individual projects. The structures determined in this thesis work were solved by **MR** and **SAD.** Detailed information about individual projects are given in the following sections.

4.2.4 Heparinase III

The **HepIII** gene (*Bt*4657) from *Bacteroidetes thetaitomicron* without the signal peptide ($HepIII^{21-666}$) was previously cloned in Dr.Cygler's lab. $HepIII^{21-666}$ was cloned in **pRL652** expression vector as a GST fusion protein with a TEV cleavage site. For protein expression, the PRL652/HepIII plasmid was transformed into chemically competent *E.coli* BL21 cells as described in section 4.2.1.4.

4.2.4.1 Protein Expression and Purification

HepIII was expressed as described in section 4.2.2. Briefly, an overnight inoculum of the *E. coli* expression strain harboring HepIII plasmid was subcultured in one liter of TB medium supplemented with ampicillin. These cells were allowed to grow at 37° C until the absorbance of the culture at 600 nm reached 1.2. The temperature was reduced to 20° C and protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After overnight growth, the cells were harvested as described in section 4.2.2 and resuspended in 30 mL of lysis buffer containing 15 mM HEPES pH 7.5, 150 mM NaCl and 5 mM dithiothreitol (DTT). The cells were lysed using a cell disruptor and the protein was purified using glutathione resin as described in section 4.2.2.1. The GST tag was cleaved at room temperature by adding TEV protease at a mass ratio of 1:50 and incubated for 8 hours. Removal of TEV protease was then carried out as described in section 4.2.2.1.

Table 4.2-3. Buffers used in HepIII project.

HepIII	
Lysis buffer	15 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT
Wash buffer	Phosphate-buffered saline (PBS) + 200 mM NaCl, PBS + 350 mM NaCl
TEV cleavage buffer and condition	15 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT, room temperature for 8 hours
SEC column and buffer	GE S200 column, 20 mM HEPES pH 7.5, 150 mM NaCl and 3 mM DTT

4.2.4.2 Methylation of HepIII

The pure fractions of HepIII were concentrated to 9 mg/mL for the methylation experiment. Twenty µl of dimethylamine borane complex (methylating agent) and 40 µl of 10% formaldehyde were added to 1 mL of HepIII solution, mixed well, and incubated at 4^{o}C for 2 hours. This step was repeated one more time. After the incubation, 10 µl of the methylating reagent was added and incubated overnight at 4^{o}C. The next day, the reaction was stopped by adding 125 µl of 1 M glycine and the methylating chemicals were removed by SEC. Methylated HepIII was passed through the GE S200 column and the protein eluted as a single peak corresponding to a monomeric molecular weight (72.5 kDa for a monomer). Those fractions representing the monomer were then analyzed by SDS-PAGE. The clean fractions were pooled and concentrated using a 10 kDa cut-off centrifugal filter for crystallization experiment.

4.2.4.3 Crystallization, X-ray diffraction and data collection of HepIII

HepIII was concentrated to 15 mg/mL and the crystallization experiments were carried out using many commercially available screens at room temperature (~20^{o}C). The initial hit conditions were optimized by the hanging drop vapor diffusion method. Both micro and macro seeding techniques were used to improve the crystal quality. The diffraction quality HepIII crystals were cryoprotected with mother liquor supplemented with 20% glycerol and flash frozen in liquid nitrogen. Diffraction data were collected at the 08ID beamline equipped with a CCD X-ray detector at the CLS (Grochulski et al., 2011). The diffraction images were processed using HKL2000 (Otwinowski and Minor, 1997).

4.2.4.4 Structure determination

The structure of HepIII was solved using the Phaser-MR program in the PHENIX package (Adams et al., 2011). The phase information was obtained from the previously solved HepIII structure in Dr.Cygler's lab. Manual inspection, model building and structure refinement were carried out using COOT and the PHENIX refinement program, respectively (Emsley and Cowtan, 2004; Adams et al., 2011). The refined structure was validated using the MolProbity server (Chen et al., 2010).

4.2.4.5 Putative active site mutants and enzyme activity

The active site mutants of HepIII, namely; Y301F, H431A, Y457F, N247A were made as described in section 4.2.1.5. These mutants were purified in the same way as the native protein and were tested for enzyme activity. The assay solution was composed of 0.15 mg/mL heparan sulfate in 20 mM HEPES pH 7.5, 150 mM NaCl and 2 mM DTT. One µg of enzyme was added directly to the assay solution. A control reaction without enzyme was used to remove the background absorbance from substrate. The plate was immediately placed on a temperature controlled SpectraMax plate reader (Molecular Devices, Sunnyvale, CA) set at 25° C. Absorbance readings at 230 nm were taken every 20 s for 30 min.

4.2.4.6 Crystallization of HepIII mutant with substrate and structure solution

One of the HepIII mutants, Y301F was purified in large scale following the protocol used for wildtype HepIII. The methylated monodisperse Y301F sample was used in various crystallization screens with substrate to obtain a complex structure. The tetrasaccharide substrates used were **heparosan dp4** (ΔUA-GlcNAc-GlcA-GlcNAc) and **heparosan dp4S** (ΔUA-GlcNS-GlcA-GlcNS). The substrates were received from the Dr. Lindhardt laboratory, United States. In one of the co-crystallization experiments, the Y301F mutant was mixed with tetrasaccharide substrates in various ratios, incubated overnight at 4° C and screened for new crystallization hits as described in section 4.2.3. Also, using the hanging drop vapor diffusion method, the protein substrate complex was screened around the condition where the wild type HepIII had crystallized. Microseeding using wild type crystals were also carried out in these screens. Crystals of Y301F alone were grown by the hanging drop vapour diffusion method in the condition giving rise to wild type HepIII crystals. For the soaking experiment, the aforementioned tetrasaccharides and a digested product of long heparin by native HepIII were used. On a cover slip, 3 µl of substrate was mixed with 3 µl of well solution and equilibrated against 500 µl of well solution. Y301F mutant crystals were soaked in the pre-equilibrated substrate solution for different time periods. For the **diffraction** experiment, substrate soaked crystals were cryo protected with mother liquor supplemented with 20% glycerol and flash cooled in liquid nitrogen. Diffraction data of substrate soaked Y301F crystal was collected at the CMCF-ID beamline at the CLS. The diffraction data were processed using HKL2000 (Otwinowski and Minor, 1997) and the structure was solved by **MR** using the MR-Phaser program available in PHENIX package (Adams et al., 2011).

4.2.4.7 Normal mode analysis (NMA)

Biologically significant conformational changes in proteins occur over a long period of time, from nanoseconds to seconds. Calculating molecular dynamics simulation for such a long timescale is computationally taxing. In such situations, NMA is a useful technique to calculate the possible large conformational changes by treating the Cα atoms in proteins as spheres and the peptide bonds connecting them as strings (Yang and Bahar, 2005; Bahar et al., 2010). In simple terms, by applying Hooke's and Newton's law of motion to a simplified model (spheres connected by strings), movements with low frequency and high amplitude are calculated. NMA applies a diagonalized Hessian matrix, in which the elements are mass weighed secondary derivatives of protein potential energy. This minimizes the complexity that would otherwise accompany the computation of values for many atoms. The resulted eigenvectors are normal modes, which describe the directions of protein motions/conformational changes, whereas the eigenvalues are the frequencies of the corresponding normal modes. The collective movements of the protein are approximately described by the superposition of several low-frequency normal modes, whereas the other normal modes mostly represent local vibrations, which do not contribute to large scale conformational changes.

In order to investigate the large-scale movements of HepIII, NMA was performed on the apo form of two proteins: One is the crystal structure of the open form of HepIII determined here and the other is the relatively closed crystal structure 4FNV homolog (Dong et al., 2012) . Calculations were done in elNemo online server, (Suhre and Sanejouand, 2004). The potential energy of the protein was described by an elastic network with a cutoff of 8 Å. Homology models of the closed state of HepIII and the open state of 4FNV were constructed and used to determine the contributions of each mode to the conformational changes. For NMA of each crystal structure, two lowest frequency modes were selected and analyzed based on their contributions. Perturbation was applied on the crystal structures to get deformed conformations along these normal modes by g_nmtraj tool in GROMACS package. The magnitude of the deformation was determined based on the distance between the deformed structure and the homology model.

4.2.5 Ulvan lyase

The **ulvan lyase** genes from marine cohabiting bacteria were identified in the laboratory of Dr.William Helbert (CNRS, France) and Dr. Ehud Banin (Bar-Ilan University, Israel). These genes were cloned without signal peptide in the **pET28a** expression vector**.** The expression plasmids containing ulvan lyase genes were obtained from them. The list of genes and the NCBI accession numbers are mentioned in table 4.2-4. In addition, a few truncated constructs of the ulvan lyase genes were made by following the LIC protocol as mentioned in section 4.2.1.1. Small scale protein expression and solubility analysis were attempted for all the received genes.

Table 4.2-4. Newly identified ulvan lyase genes.

Strains	Short ulvan lyase	Long ulvan lyase
Alteromonas sp. LOR	LOR_107 (59.6 KDa) *WP_052010193.1*	LOR_61 (110.9 KDa) *WP_032096165.1*
Alteromonas sp. LOR	LOR_29 (53.7 KDa) *WP_052010178.1*	
Pseudoalteromonas sp. PLSV	PLSV_3875 (59.6 KDa) *WP_033186987.1*	PLSV_3925 (111.4 KDa) *WP_033186955.1*
Pseudoalteromonas sp. PLSV	PLSV_3936 (50 KDa) *WP_033186995.1*	
N. ulvanivorans NLR	NLR_48 (33.4 KDa) *WP_036579443.1*	NLR_42 (59 KDa) *WP_036579437.1*
N. ulvanivorans NLR	NLR_492 (50 KDa) *WP_036580476.1*	

4.2.5.1. Production and Purification of wildtype LOR107 and PLSV3936

LOR_107 and *PLSV_3936* genes have several codons that are rare in *Escherichia coli* so BL21(DE3)-RIPL cells were used for protein production. The protein was expressed as described in section 4.2.2. Briefly, an overnight inoculum of the expression strain containing the plasmid was sub-cultured into one liter of TB medium supplemented with 50 µg kanamycin and 50 µg chloramphenicol. Cells were allowed to grow at 37° C until the absorbance at 600 nm reached 1.5, then the temperature was lowered (reduced to 18° C and 20° C for LOR107 and PLSV3936 respectively) and protein expression was induced with 0.5 mM IPTG. After overnight growth the cells were harvested and lysed using a cell disruptor. LOR107 and PLSV3936 were cloned with an N- terminal 6X His tag and purified by metal affinity chromatography using cobalt talon resin

following the protocol delineated in section 4.2.2.1. Buffers used in the purification are listed in table 4.2-5. The LOR107 and PLSV3936 eluted with 40 mM imidazole were concentrated using a 10 kDa cut-off centrifugal filter and injected onto a GE S200 size exclusion column. The protein fractions eluted in a single peak with the molecular weight corresponding to that of a monomer. These peak fractions were pooled together and concentrated. In the case of LOR107 the protein was concentrated to 17 mg/mL and PLSV3936 was concentrated to 29 mg/mL. This concentrated protein solution was then screened for crystallization. Problems were encountered in the expression of LOR107. The protein did not express when grown in large scale (1 litre TB). To overcome this difficulty, the LOR107 protein was expressed in five aliquots of 200 mL TB media. All the mutants of LOR107 were expressed in this way.

Table 4.2-5 Buffers used in LOR107 andPLSV3936 purification

LOR107 & PLSV3936	
Lysis buffer	20 mM Tris pH8.0, 200 mM NaCl, 5% glycerol and 5 mM imidazole
Wash buffer	Lysis buffer +20 mM imidazole Lysis buffer +40 mM imidazole
SEC column and buffer	Bio-Rad SEC650 and GE S200, 20 mM HEPES pH 7.5, 150 mM NaCl

4.2.5.2. Production and purification of SeMet incorporated LOR107 and PLSV3936

Selenomethionine was incorporated into LOR107 and PLSV3936 by inhibiting the methionine biosynthesis pathway (Walden, 2010). Briefly, an overnight inoculum in 100 mL LB media was grown at 37° C. The cells were then spun down and the pellet was resuspended in M9 minimal media. The resuspended cells were used to subculture 1 L of M9 minimal media supplemented with 50 µg kanamycin and 50 µg chloramphenicol. Cells continued to grow at 37^{o}C until the optical density at 600nm (OD_{600}) reached ~1 (5-6 h). Then 100 mg of hydrophobic amino acids lysine, phenylalanine, threonine, 50 mg of isoleucine, leucine, valine and 60 mg of L-selenomethionine were added. After a 15 min incubation, protein expression was induced by the addition of 0.5 mM IPTG and the cells continued to grow at 20^{o}C for 16 hours. Next, cells were spun down and the SeMet incorporated protein was purified in the same way as native protein

(except 2 mM DTT was included in the lysis buffer and purification buffer in the SeMet incorporated protein purification). After passing through a GE S200 column, the LOR107 and PLSV3936 fractions corresponding to the monomeric form of protein were concentrated to 20 mg/mL and 30 mg/mL, respectively. These concentrated samples were used for hanging drop crystallization experiments.

4.2.5.3. Site directed mutagenesis and enzyme activity assay

Several putative active site mutants were made for LOR107 and PLSV3936 by following the protocol described in section 4.2.2.5. The first set of mutants made for **LOR107** are as follows: H146A, H167A, R259A, T242A, Y243F, N263A, Y303A, R320A, Y330F, R259N and R259L. The second set of LOR107 mutants: H146N, H167N, R259N, R259L, R320N and R320L were made specifically to obtain a complex structure.

The putative active site mutants made in **PLSV3936** are H123N, H143N, Y188F, R204N and H264N. The list of primers used in the site directed mutagenesis PCR to make the aforementioned mutants are listed in section 4.3.1.4.

The enzyme activity of wild-type and various mutants of **LOR107** and **PLSV3936** were measured by monitoring the absorption at 232 nm on a SpectraMax plate reader set at 25°C. The assay solution was composed of 0.5 mg/mL ulvan polysaccharide in 20 mM HEPES pH 7.5 and 150 mM NaCl. One µg of the enzyme was added directly to the assay solution with constant stirring. The absorbance of the substrate alone was used as a control.

4.2.5.4. Crystallization and diffraction of wild type and SeMet incorporated LOR107

Crystallization screening of wild type LOR107 was carried out at room temperature (~20^{o}C) using various commercial screens as described in section 4.2.3. The initial hit condition was optimized by the hanging drop vapor diffusion method. For data collection, the LOR107 crystals were cryoprotected by briefly immersing in the mother liquor and were flash cooled in liquid nitrogen. Diffraction data were collected at the 08ID beamline at the CLS equipped with Rayonix MX300 CCD detector (Grochulski et al., 2011). The crystal to detector distance was set to 250 mm. A total of 360 degrees with 0.7^{o} oscillation was collected. This recorded 520 diffraction frames. The diffraction data was processed with the XDS program (Kabsch, 2010). The crystallization of SeMet-LOR107 was attempted by hanging drop vapor diffusion method through

systematic grid screening around the condition where wild type LOR107 had crystallized. The best crystals of SeMet-LOR107 were produced through several rounds of micro seeding. The initial round of seeding was done using crushed crystals of wildtype LOR107 and the following several rounds of seeding were done using microcrystals of SeMet-LOR107. For data collection, the SeMet-LOR107crystals were briefly transferred to mother liquor and flash cooled in liquid nitrogen. Diffraction data were collected at the 08ID beamline (Grochulski et al., 2011) at the CLS. The SeMet-LOR107 crystals were scanned for Se absorption peak and the diffraction data was collected at the Se absorption edge wavelength. The crystal to detector distance was set to 320 mm. A total of 360 degrees with a 0.75^{o} oscillation was collected. This recorded 480 diffraction frames. Diffraction data was processed using HKL 3000 (Minor et al., 2006) by treating the Friedel pairs as different reflections.

4.2.5.5. Crystallization and diffraction of wild type and SeMet incorporated PLSV3936

Like LOR107, crystallization screening of wild type PLSV3936 was carried out at room temperature (~20^{o}C) using various commercial screens. Initial crystallization hit conditions of wild type PLSV3936 were optimized by the hanging drop vapor diffusion method through systematic grid screening. The crystallization of SeMet-PLSV3936 was achieved by the hanging drop vapor diffusion method through systematic grid screening around the wild type crystallization condition. For data collection, both wild type and SeMet crystals were cryo protected with well solution supplemented with 30% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data of wild type and SeMet crystals were collected at the 08B1 beamline at the CLS (Fodje et al., 2014) equipped with a Rayonix MX300HE detector. For the wildtype crystal, a total of 360 diffraction frames with 0.5^{o} oscillation were collected. The crystal to detector distance was set to 190 mm. Diffraction data was processed with XDS (Kabsch, 2010). The SeMet crystals were scanned for Se absorption peak and the diffraction data was collected at the Se absorption edge wavelength. The crystal to detector distance was set to 250 mm. A total of 514 frames with 0.7^{o} oscillation was collected. Diffraction data was processed using HKL 3000 (Minor et al., 2006) by treating the Friedel pairs as different reflections.

4.2.5.6. Structure determination and refinement of LOR107 and PLSV3936

The structure of **LOR107** was solved from the SeMet data using the automated structure solving pipeline **CRANK2** (Skubák and Pannu, 2013). Structure of **PLSV3936** was solved using

autosol program (Terwilliger et al., 2009) in PHENIX using the data from the SeMet crystal. Both programs have done all the steps including Se substructure identification, density modification and hand correction. This led to the generation of an initial model. Both structures were refined with the PHENIX software (Adams et al., 2011) and manual rebuilding and solvent placement was conducted with the COOT program (Emsley and Cowtan, 2004). The stereochemistry of both models were validated with MolProbity (Chen et al., 2010).

4.2.5.7. Purification and crystallization of protein-ligand complex

The tetrasaccharide substrates used in the co-crystallization and soaking experiments were obtained from Dr. William Helbert (CNRS, France). They were prepared by enzymatic digestion of ulvan polysaccharide with LOR107 ulvan lyase (PL24) from *Alteromonas* LOR *sp*. (Kopel et al., 2016) and separated by HPLC chromatography. Tetrasaccharide I (DP4-1) shown in **Figure 4.2:1** contains a mixture of ΔUA-Rha3S-GlcA-Rha3Sα/β and ΔUA-Rha3S-IdoA-Rha3Sα/β with some traces of hexasaccharides. Tetrasaccharide II (DP4-2) is a pure sample of ΔUA-Rha3S-Xyl-Rha3Sα/β.

To obtain the complex structure of **LOR107** with substrate, the following active site mutants, H146N, H167N, Y243F, R259N, R259L and R320N were expressed and purified in the same way as native LOR107. The purified mutants were passed through the GE S200 column to separate aggregates and different oligomeric states. The monomeric fractions of LOR107 mutants (namely R259N, R259L, Y243F and R320N) were concentrated to 22 mg/mL and screened for crystallization. In the **co-crystallization** experiments, the R259N mutant and dp4-1 tetrasaccharide were mixed at various ratios and incubated for a few hours at 4^{o}C. The protein-ligand mixture was screened for crystallization as described in section 4.2.2.2. For the **soaking** experiment, crystals of mutant (R259N, R320N and Y243F) were grown by a grid screen around the condition where the native protein crystallized. Several rounds of microseeding were also carried out to obtain mutant crystals of LOR107.

In the **PLSV3936** project, the mutants H123N, Y188F and H264N were purified in the same way as the native protein. The monomeric fractions of H123N and H264N mutants from the SEC experiment were concentrated to approximately 29 mg/mL and Y188F mutant to 17 mg/mL. All the three mutants were screened for crystallization by the hanging drop vapor diffusion method around the conditions where native crystals grew.

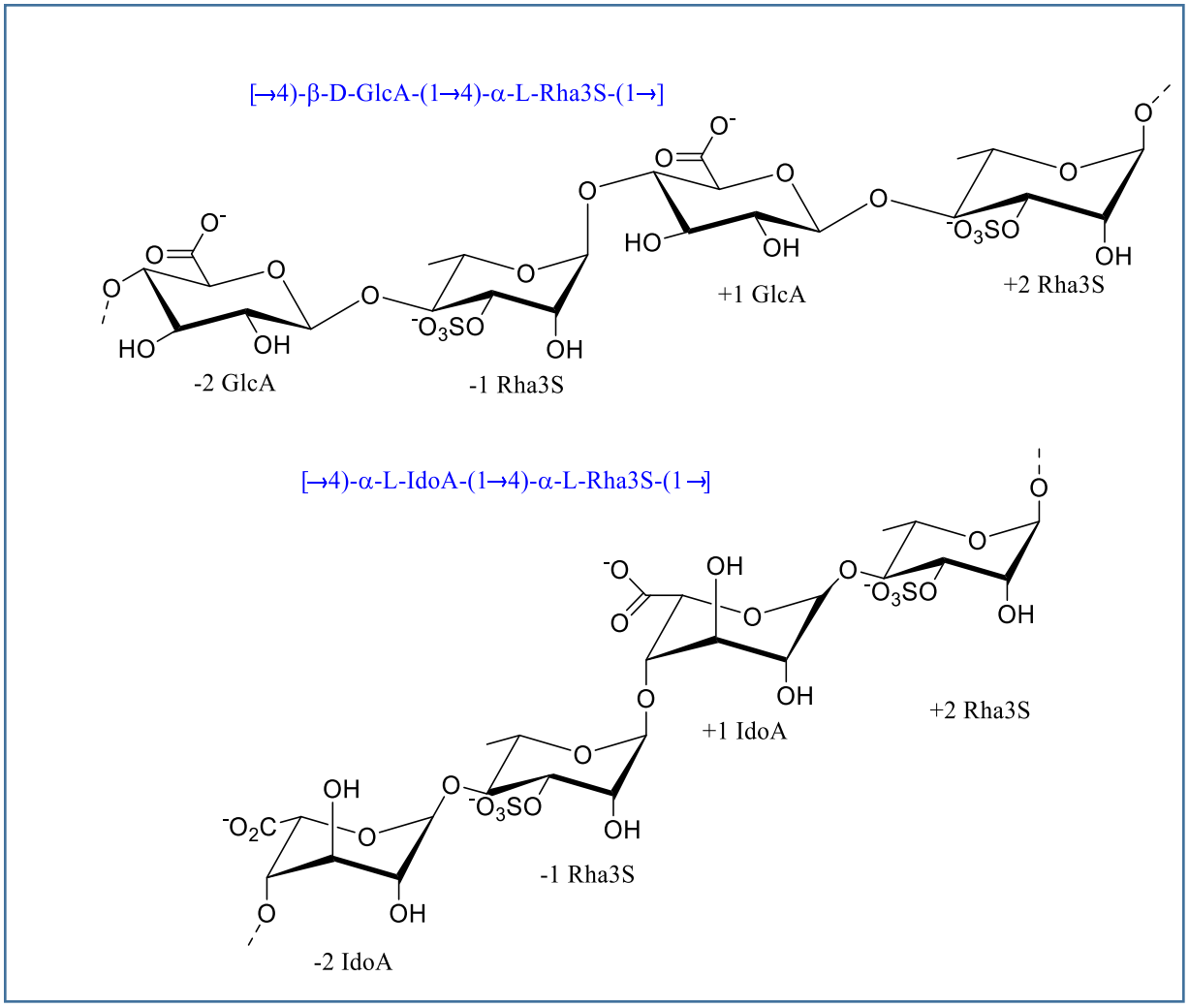

Figure 4.2:1 Schema of ulvan tetrasaccharide

4.2.5.8. Soaking experiment and data collection

For soaking experiments, 3 µl of tetrasaccharide substrate was mixed with 3 µl of well solution in a cover slip and equilibrated against 1 mL of well solution. This soaking experiment was done with both dp4-1 and dp4-2 substrates. Also, soaking experiments were carried out at various incubation times, ranging from 5 minutes to 8 hours in length.

Crystals of **LOR107** single mutants, R259N, Y243F and R320N, were soaked for ~5 hours in the two different tetrasaccharide substrates. After soaking, the crystals were flash cooled in liquid nitrogen. Diffraction data of the **R259N** crystals were collected at the 08B1 beamline (Fodje et al., 2014), at the CLS. The crystal to detector distance was set to 300 mm. A total of 360 frames

were collected with 1^{o} oscillation. Data was processed with the XDS software (Kabsch, 2010). For R320N substrate soaked crystals, the diffraction data of the crystals were collected at the 08ID beamline (Grochulski et al., 2011), at the CLS.

In **PLSV3936**, the H123N mutant crystals were soaked in the substrate solution at various time points. After soaking, the crystals were cryo protected with mother liquor containing 30% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data were collected at the 08B1 beamline at the CLS (Fodje et al., 2014) and processed with XDS (Kabsch, 2010). The crystal to detector distance was set to 210 mm. A total of 514 frames were collected with 0.7^{o} oscillation.

4.2.5.9. Complex structure solution and ligand modelling

The mutant crystals were isomorphous to the native crystal, and one round of rigid body refinement was done on the complex dataset with the PHENIX refinement program. The coordinates and the restraint file for the tetrasaccharide were created using Jligand program (Lebedev et al., 2012) available in the CCP4 package and elbow (Moriarty et al., 2009) from the PHENIX package. The complex structure of both LOR107 and PLSV3936 mutants were refined using the PHENIX refinement program. Modelling of ligand into the electron density, structural inspection and other aspects of model building were carried out using COOT (Emsley and Cowtan, 2004). The stereochemistry of these structures was validated using MolProbity (Chen et al., 2010).

4.2.6 NLR48

The construct NLR48$^{(26\text{-}302)}$ lacking the signal peptide was cloned into expression vectors with different tags to improve solubility and to aid in purification. The NLR48$^{(26\text{-}302)}$ fusion constructs were made by following the LIC protocol. GST-NLR48$^{(26\text{-}302)}$ fusion was made by cloning into the **pRL652** vector; His-MBP-NLR48$^{(26\text{-}302)}$ fusion was made by cloning into the **pMCSG9** vector. Based on secondary structure predictions, a shorter construct lacking a few residues at the N-terminal unstructured region of NLR48$^{(53\text{-}302)}$ was cloned into the **pMCSG7** vector.

4.2.6.1. Expression and purification of wild type NLR48

His-NLR48$^{(26\text{-}302)}$, His-MBP-NLR48 $^{(26\text{-}302)}$) GST-NLR48$^{(26\text{-}302)}$ and His-NLR48$^{(53\text{-}302)}$were expressed as described in section 4.2.2. The 6X His and His-MBP tag constructs were purified using Ni-NTA metal affinity chromatography as described in section 4.2.2.1. Although various

NLR48 constructs were made, only GST-NLR48$^{(26-302)}$ was used for structural studies and active site mutant construction. Briefly, GST-NLR48$^{(26-302)}$ was expressed and purified as follows: The GST-NLR48$^{(26-302)}$ plasmid was transformed into BL21(DE3) expression cells. The protein was expressed with 0.5mM IPTG at 18^{o}C. The GST-NLR48$^{(26-302)}$ fusion protein was sensitive to higher temperature, so all the purification steps were carried out at 4^{o}C. After overnight growth, the cells were harvested and resuspended in 40 ml of lysis buffer containing 20 mM Tris pH8.0, 400 mM NaCl, and 5% glycerol. The cells were then disrupted by sonication (7 seconds on, 7 seconds off for 4 minutes, output level 5) and the protein was purified using glutathione resin (as described in section 4.2.2.1). The GST tag was cleaved at 4°C by TEV protease added at the mass ratio of 1:50 over the course of ~14 hours. The TEV protease was removed as described in section 4.2.2.1. Purified NLR48$^{(26-302)}$ was then injected onto a Bio-Rad SEC650 column to check its oligomeric state. The peak fractions corresponding to the monomeric molecular weight of the protein were concentrated to 16 mg/ml using a 30 kDa cut-off centrifugal filter. This concentrated protein was then used for crystallization experiments.

Table 4.2-6. Buffers used in NLR48 purification.

NLR48 (all purification steps were carried out at 4^{o}C)	
Lysis buffer	20 mM Tris pH8.0, 400 mM NaCl, and 5% glycerol
Wash buffer	200 mM NaCl+lysis buffer
TEV cleavage buffer and condition	20 mM Tris pH8.0, 400 mM NaCl, and 5% glycerol, ~12 hours at 4^{o}C
SEC column and buffer	Bio-Rad SEC650, 20 mM Tris pH8.0, 150 mM NaCl

4.2.6.2. Selenomethionine incorporation of NLR48

Expression plasmid carrying the GST-NLR48$^{(26-302)}$ fusion construct was transformed into **methionine auxotroph** cells B834 (DE3). An overnight inoculum grown in 50 mL M9 medium supplemented with methionine was subcultured into 1 L M9 medium containing methionine. For selenomethionine incorporation, cells were grown at 37˚C until OD_{600} value reaches 1 and cells were spun down in sterile centrifuge bottles at 3000 x g using beckman JLA9.1 rotor for 20 minutes. Cell pellets were resuspended in 1 L of M9 medium and the culture was incubated at 37˚C for another 4 hours to deplete the remaining methionine. After depletion, 50 mg/L of

selenomethionine was added to the cell culture 30 minutes prior to the addition of IPTG. The cells were further grown at 18° C. After ~16 hours of growth, the cells were spun down and selenomethionine incorporated protein was purified following the same protocol used for native enzyme purification.

4.2.6.3. Crystallization and diffraction experiment of wild type and SeMet-NLR48$^{(26-302)}$

The GST tag cleaved NLR48$^{(26-302)}$ was concentrated to 16 mg/mL and screened for crystallization by following the protocol described in 4.1.3. The crystallization plates were maintained at 15° C and 9° C. Both **wildtype** and **SeMet-**NLR48$^{(26-302)}$ crystals were improved by the hanging drop vapor diffusion method through grid screening. Microseeding using wild type NLR48$^{(26-302)}$ crystals was carried out to improve the quality of SeMet-NLR48$^{(26-302)}$ crystals. For diffraction data collection, the SeMet crystals were cryoprotected with 30% ethylene glycol and flash frozen in liquid nitrogen. Diffraction data was collected at the 08BM beamline at the CLS. The crystal to detector distance was set to 250 mm and a total of 370 frames were collected with 1^{o} oscillation. Data was processed with XDS package (Kabsch, 2010).

4.2.6.4. Active site mutants- enzyme activity and crystallization

Putative active site single mutants Q137A, K139M, Y258F, R94N and R193N were made using the GST-NLR48$^{(26-302)}$ fusion plasmid as a template. All clones were confirmed by sequencing. The active site mutants were expressed and purified in the same way as the wildtype enzyme. Enzyme activity of wild type and single mutants were measured using the protocol described for other ulvan lyase (section 4.2.5.3).

The active site mutants K139M and Y258F were concentrated to 17 mg/mL and screened for crystallization. Crystals were grown by the handing drop vapor diffusion method at both 15° C and 9° C. Crystal growth was promoted by microseeding using wildtype crystals. The substrate soaking solution was made as described in section 4.2.5.8. The mutant crystals were soaked in the dp4-1substrate solution for ~2 hours and cryo-protected with well solution containing 30% ethylene glycol followed by flash cooling in liquid N_2. Diffraction data were collected at the 08ID beamline (Grochulski et al., 2011) equipped with Pilatus3 S 6M detector at the CLS. The crystal to detector distance was set to 300 mm and 1350 frames were collected with 0.2^{o} oscillation. The mutant crystal was isomorphous to the wild type NLR48 crystals. One round of rigid body

refinement was done on the complex dataset. The refinement was performed with the PHENIX refinement program. Manual building of the substrate into the positive density, structure inspection and rebuilding were done using COOT (Emsley and Cowtan, 2004). Further refinement was done using the PHENIX refinement program (Adams et al., 2011). The complex structure was evaluated for stereochemistry using MolProbity (Chen et al., 2010).

4.3 . List of Materials

4.3.1 List of Primers used in various projects

4.3.1.1. HEPIII mutants

N247A	5’ CTCTGATCAGGGTGCTCACTTGTTGTTC 3’ 5’ GAACAACAAGTGAGCACCCTGATCAGAG 3’
Y457F	5’ GTTCGTATGTGTTTGCCGGTGAAGG 3’ 5’ CCTTCACCGGCAAACACATACGAAC 3’
H431A	5’ CCGAAAGGTTTCTGGGCCTGTCAGCCGGATAAC 3’ 5’ GTTATCCGGCTGACAGGCCCAGAAACCTTTCGG 3’
Y303F	5' CTTGACCCGCATTTTCATCTTGCTGC 3' 5' GCAGCAAGATGAAAATGCGGGTCAAG 3'

4.3.1.2. LOR107

H146A	5’ GTAGGTGAATCGGCCAACACCATTG 3’ 5’ CAATGGTGTTGGCCGATTCACCTAC 3’
H167A	5’ GTTTACGACATGGCCGCCTATGTTG 3’ 5’ CAACATAGGCGGCCATGTCGTAAAC 3’
H167N	5’ GTTTACGACATGAACGCCTATGTTG 3’ 5’ CAACATAGGCGTTCATGTCGTAAAC 3’
H167F	5’ GTTTACGACATGTTCGCCTATGTTG 3’ 5’ CAACATAGGCGAACATGTCGTAAAC 3’
T242A	5’ TTTTCTGCGCTAGCCTATCCTAAAT 3’ 5’ ATTTAGGATAGGCTAGCGCAGAAAA 3’
Y243F	5’ CTGCGCTAACCTTTCCTAAATTTTAC 3’ 5’ GTAAAATTTAGGAAAGGTTAGCGCAG 3’
R259A	5’ GCTTCACTATATGGCCTGGGGTGGCAA 3’ 5’ TTGCCACCCCAGGCCATATAGTGAAGC 3’
R259N	5’ GCTTCACTATATGAACTGGGGTGGCAA 3’ 5’ TTGCCACCCCAGTTCATATAGTGAAGC 3’
R259L	5’ GCTTCACTATATGCTCTGGGGTGGCAA 3’ 5’ TTGCCACCCCAGAGCATATAGTGAAGC 3’
N263A	5’ ATGCGCTGGGGTGGCGCTAACAACGGCGCTTATTACTT 3’ 5’ AAGTAATAAGCGCCGTTGTTAGCGCCACCCCAGCGCAT 3’

Y303F	5’ CAATTGGGGACTGTTTGGCCAAATGAAG 3’
	5’ CTTCATTTGGCCAAACAGTCCCCAATTG 3’
R320A	5’ GTAGGGTTTCAACAGGCTAGCGCAAATAATG 3’
	5’ CATTATTTGCGCTAGCCTGTTGAAACCCTAC 3’
R320N	5’ CGTAGGGTTTCAACAGAATAGCGCAAATAATG 3’
	5’ CATTATTTGCGCTATTCTGTTGAAACCCTACG 3’
Y330F	5’ GCAAATAATGATGACCGTTTTAAATTTCAAAATGGTGTTTATTATGCC 3’
	5’ GGCATAATAAACACCATTTTGAAATTTAAAACGGTCATCATTATTTGC 3’

4.3.1.3. PLSV3936

H123F	5’ GAAATTCGATAATTTTGGTAAGCCAACC 3’
	5’ GGTTGGCTTACCAAAATTATCGAATTTC 3’
H123N	5’ GAAATTCGATAATAATGGTAAGCCAACC 3’
	5’ GGTTGGCTTACCATTATTATCGAATTTC 3’
H143F	5’ TTTTTACGGCGGATTTGGCGGTCAGGCA 3’
	5’ TGCCTGACCGCCAAATCCGCCGTAAAAA 3’
H143N	5’ TTTTTACGGCGGAAATGGCGGTCAGG 3’
	5’ CCTGACCGCCATTTCCGCCGTAAAAA 3’
R204L	5’ ATTTATTTTTCCTCCACGGTGCTCA 3’
	5’ TGAGCACCGTGGAGGAAAAATAAAT 3’
R204N	5’ ATTTATTTTTCAACCACGGTGCTCA 3’
	5’ TGAGCACCGTGGTTGAAAAATAAAT 3’
H264F	5’ CCTACGACTATTTCGTTTGCTGGGA 3’
	5’ TCCCAGCAAACGAAATAGTCGTAGG 3’
H264N	5’ CCTACGACTATAACGTTTGCTGGGA 3’
	5’ TCCCAGCAAACGTTATAGTCGTAGG 3’
Y188F	5’ CCATTTGGTACCTTTAACCAAGCAATC 3’
	5’ GATTGCTTGGTTAAAGGTACCAAATGG 3’

4.3.1.4. NLR48

Q137A	5’ GCTACTTTATGGCGGCTAAAGGCAAA 3’
	5’ TTTGCCTTTAGCCGCCATAAAGTAGC 3’
K139H	5’ CTTTATGCAGGCTCACGGCAAACACACAG 3’
	5’ CTGTGTGTTTGCCGTGAGCCTGCATAAAG 3’
K139M	5’ CTTTATGCAGGCTATGGGCAAACACAC 3’
	5’ GTGTGTTTGCCCATAGCCTGCATAAAG 3’
K139A	5’ TTTATGCAGGCTGCAGGCAAACACA 3’
	5’ TGTGTTTGCCTGCAGCCTGCATAAA 3’
Y258F	5’ AGATATGGTGCCTTCCGTGTTAAAG 3’
	5’ CTTTAACACGGAAGGCACCATATCT 3’
R94N	5’ CATTACAGCCTAACATAGAACGTTCC 3’
	5’ GGAACGTTCTATGTTAGGCTGTAATG 3’
R94A	5’ CATTACAGCCTGCGATAGAACGTTC 3’
	5’ GAACGTTCTATCGCAGGCTGTAATG 3’

R193N	5’ CAGGTTCTGCTGGAAATACTGAAGTTTTTC 3’ 5’ GAAAAACTTCAGTATTTCCAGCAGAACCTG 3’
R193A	5’ GGTTCTGCTGGAGCAACTGAAGTTTTTC 3’ 5’ GAAAAACTTCAGTTGCTCCAGCAGAACC 3’

4.3.2 List of Chemicals

100 mM dCTP and 100 mM dGTP	NEB, N0446S
100×BSA	NEB
30% Acrylamide/Bisarcylamide(29:1)	Bioshop, ACR009.500
Ammonium persulfate (APS)	BioShop, AMP001.25
Bis-Tris	BioShop, BST602.250
Borane dimethylamine complex	Sigma-Aldrich, 180238-5G
Bromophenol Blue	Sigma-Aldrich, 114413-5G
Coomassie Brilliant Blue R-250	AMRESCO, (0427-25G)
Dithiothreitol (DTT)	Thermo Fisher Scientific, BP17225
DNA loading buffer	Norgen, 28129
*Dpn*I	NEB, R0176L
Enrich SEC650	Bio-Rad, 780-1650
Formaldehyde	Polysciences, Inc, 04018
Gelgreen nucleic acid gel stain	Biotium, 41005
Glutathione superflow resin, 635608	Clontech
Glycerol	Thermo Fisher Scientific, G31-4
Glycine	Thermo Fisher Scientific, BP381-5
HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	BioShop, HEP001.1
Hitrap chelating column	GE Healthcare Life Science, 17-0409-03
Imidazole	BioShop, IMD508.1
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Thermo Fisher Scientific, BP1755100

KOD hot start DNA polymerase kit	Novagen, 71086
L-Glutathione (GSH)	BioShop, GTH003.25
MPD (Methyl-pentane-diol)	Hampton Research, Aliso Viejo, CA
Nickel (II) sulfate hexahydrate	Sigma-Aldrich, 227676-500G
Ni-NTA agarose	Qiagen, 142338540
Poly-ethylene glycol (PEG) Molecular weight 400, 3500,6000, 8000	Hampton Research, Aliso Viejo, CA
Polyethylene glycol 3350 (PEG3350)	Sigma-Aldrich, P4338-1KG
Potassium phosphate dibasic (K_2HPO_4)	BioShop, PPD303.1
Potassium phosphate monobasic (KH_2PO_4)	Thermo Fisher Scientific, P285-500
Protein marker	Bio-Rad, 161-0377
QIAprep Spin Miniprep Kit (250)	Qiagen, 27106
QIAquick Gel Extraction Kit (250)	Qiagen, 28706
Sodium chloride (NaCl)	BioShop, SOD002.10
Sodium dodecyl sulfate (SDS)	BioShop, SDS001.1
Superdex 200 10/300	GE Healthcare Life Science, 2899094
T4 DNA ligase and corresponding buffer	NEB, M0202S
T4 DNA polymerase and corresponding buffer	NEB, M0203L
TEMED (*N*,*N*,*N*′,*N*′-Tetramethylethane)	BioShop , TEM001.25
Tryptone	Thermo Fisher Scientific, BP1421-2
Yeast Extract	BioShop, YEX401.205

5. Structure and Substrate Specificity of HepIII from *Bacteroides Thetaiotaomicron* (Bt4657)

The objective of this section is to describe the substrate specificity and catalytic mechanism of HepIII (PL12 family) from *Bacteroides Thetaiotaomicron.* To accomplish this, we first solved the crystal structure of wildtype HepIII. Second, we identified the putative active site residues through a structural comparison with a homolog and confirmed these residues to be catalytic by enzyme activity assays. At the same time, in a collaboration with Dr. Robert Linhardt (Rensselaer Polytechnique institute, USA) we studied the substrate specificity of HepIII using natural and chemically modified heparin/heparan sulfate polysaccharides. Third, we attempted to obtain the structure of the HepIII-substrate complex. Despite extensive efforts, we did not succeed in obtaining crystals of the HepIII-substrate complex or HepIII-product complex with either wildtype or active site mutants. However, structural comparison with HepIII homologs revealed structural flexibility in the domain arrangement, indicating an open-closed states of the substrate-binding site of HepIII. Finally, in a collaboration with Dr. D. Peter Tieleman's group at the University of Calgary, we performed Normal Mode Analysis (NMA), and modelled a possible large-scale motion of the HepIII structure during its catalytic cycle.

5.1 Expression, purification and methylation of HepIII

Expression of HepIII$^{21-666}$ was carried out in the *E. coli* BL21 cell line. The protein was purified as described in section 4.2.4.1. To increase the crystallization propensity, HepIII was methylated and the methylating reagents were removed by SEC. The purified fractions and the SEC fractions of HepIII were analyzed in a 10% SDS-PAGE **(Figure 5.1A,C)** The SEC fractions of HepIII corresponding to the monomeric size of the protein were concentrated and used for crystallization experiments.

5.2 Crystallization, X-Ray diffraction and data collection of HepIII

Methylated HepIII was concentrated to 15 mg/ml in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl and 3 mM DTT. Screening for crystallization and optimization was carried out as described in section 4.2.4.3. Diffraction quality crystals were produced in 25% (W/V) PEG 3350, 100 mM Tris pH 8.5, 200 mM $MgCl_2$ and 500 mM NDSB-201. These crystals were morphologically different from the previously obtained crystals, which were used to solve the structure of HepIII. The new crystals described here are referred to as form 2. For data collection the crystals were transferred to a cryo solution containing 20% glycerol in mother liquor and were flash cooled with liquid nitrogen. Diffraction data were collected at the 08ID beamline (Grochulski et al., 2011) at the CLS and processed with HKL2000 (Otwinowski and Minor, 1997). These crystals belong to space group $P2_1$ with unit cell parameters a=66.4, b=80.4, c=83.5 Å, $\alpha=\gamma=90$, $\beta=104.1^{o}$ and diffracted to 2.4 Å resolution. Data collection statistics are summarized in Table 5.1.

5.3 Structure determination and refinement

The structure was solved by molecular replacement with the program Phaser (McCoy et al., 2007) using the previously solved HepIII from our lab as a search model. Subsequent model building and refinement were performed using ARP/wARP (Langer et al., 2008), COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2011). One molecule of HepIII is present in the asymmetric unit. The final model contains residues 22-666, one Mg^{2+} ion and 121 water molecules. The refinement of the model converged to Rwork/Rfree of 0.215/0.256.

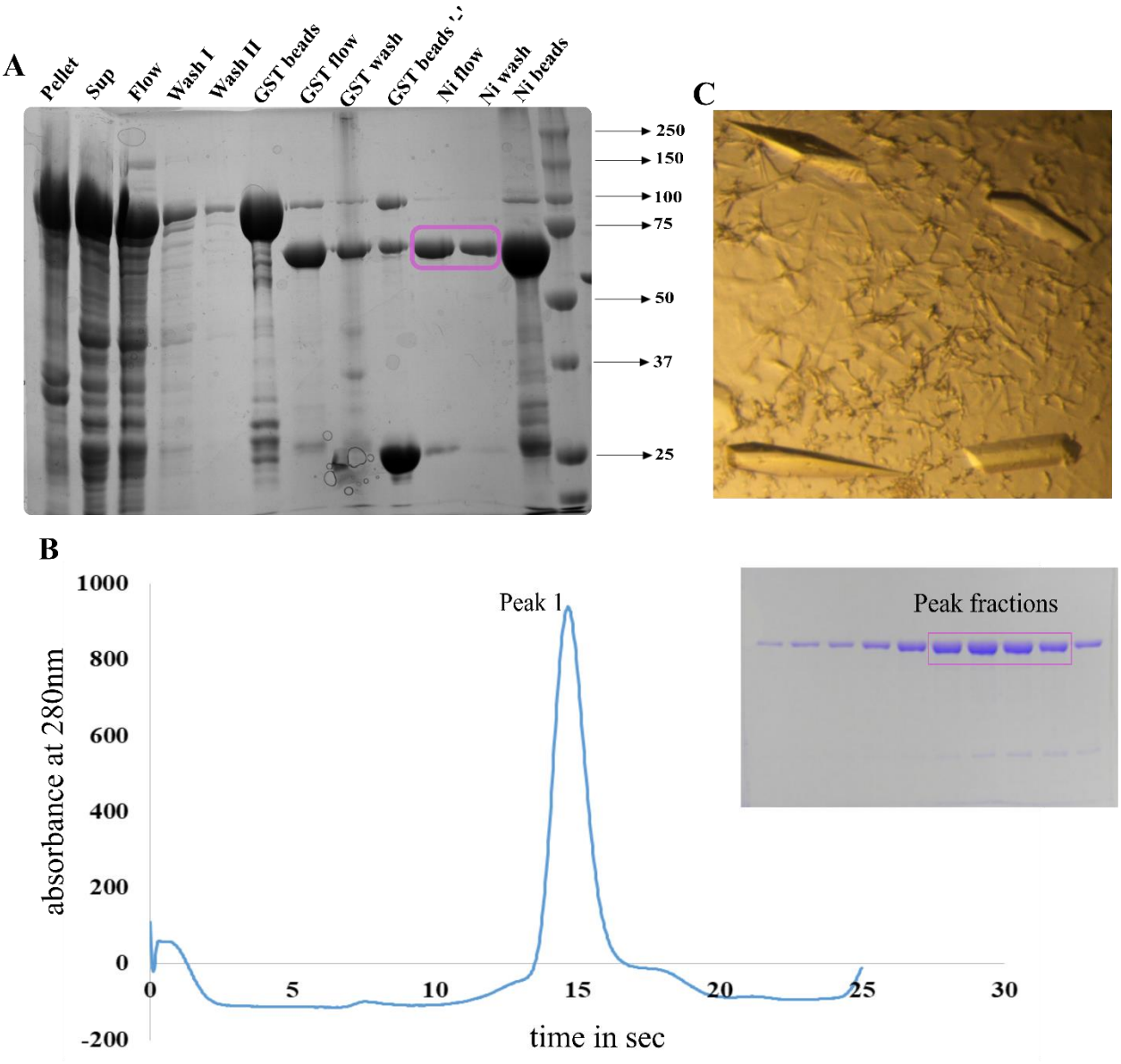

Figure 5:1Purification and crystallization of wildtype HepIII$^{(22-666)}$.

A) Affinity purification of the GST-HepIII using Glutathione agarose resin, cleavage of the GST tag using His-TEV protease and the follow up purification using Ni-NTA resin. Protein used for methylation after the GST cleavage and TEV removal is highlighted in box. **B)** SEC on the S200 column showing monomeric protein and the SDS-PAGE showing the purity of the peal fractions. **C)** Crystals of methylated HepIII$^{(22-666)}$ obtained from 25% (W/V) PEG 3350, 100 mM Tris pH 8.5, 200 mM $MgCl_2$ and 500 mM NDSB-201.

5.4 Crystal Structure of HepIII

The overall structure of HepIII consists of two domains: an N-terminal α-helical and a C-terminal β-sheet domain **(Figure 5.2A)**. The N-terminal domain is comprised of residues 21-387 and is folded into 16 α-helices. The majority of the domain is formed by ten α-helices forming five hairpin pairs arranged into an incomplete toroid fold, designated $(\alpha/\alpha)_5$. The α-helices forming five hairpins are HP1 (α6-α7), HP2 (α9-α10), HP3 (α11-α12), HP4 (α13-α14) and HP5 (α15-α16). The helices α6, α9, α11, α13 and α15 form the inner surface of this partial toroid and helices, α7, α10, α12, α14 and α16 form an outer wall of the toroid. The loops within the hairpins are rather short while the loops connecting the hairpins are significantly longer. The N-terminal segment of ~110 residues form three α-helices (α1, α2, α3) along the side of the hairpin HP1 and on the top of HP2 and HP3; this is followed by a long linker of 18 residues connecting to an α-helix (α4), β-hairpin and α-helix (α5) **(Figure 5.2B)**.

The C-terminal domain is comprised of residues 397- 666 and is folded into three twisted, antiparallel β-sheets. The top β-sheet is close to the N-terminal domain and contains 8 β-strands (β1-4, β5, β10, β14 and β15). A helix (H1) is inserted between β4 and β5. The middle layer also contains 8 β-strands (β6-9, β11-13 and β16). The bottom layer contains only 3 β-strands (β17-19) **(Figure 5.2D)**. A nine-residues linker connects the N terminal α-helical and C terminal β-sheet domains.

A metal ion was identified beneath the N-terminal domain and at the top β-sheet of the C-terminal domain. Based on the coordination geometry and the liganding residue distance, the metal ion was identified as Mg^{2+}. This Mg^{2+} ion is coordinated in an octahedral fashion by Gln^{433}, Asp^{451}, His^{476} and three water molecules. The ion-liganding atom distances vary between 2.16 and 2.29 Å. The identity of this Mg^{2+} ion was further supported during refinement, which produced no negative density at this site and gave a similar B-factor for Mg^{2+} to those of the liganding atoms (Zheng et al, 2008).

The methylation of surface-exposed lysine residues was essential for the crystallization of full length HepIII $^{(22\text{-}666)}$. Inspection of the HepIII electron density indicated that many lysine sidechains are partially disordered and the identification of methylated lysines from the electron density map is inconclusive. The coordinates were deposited in the PDB and the accession number is **5JMD**.

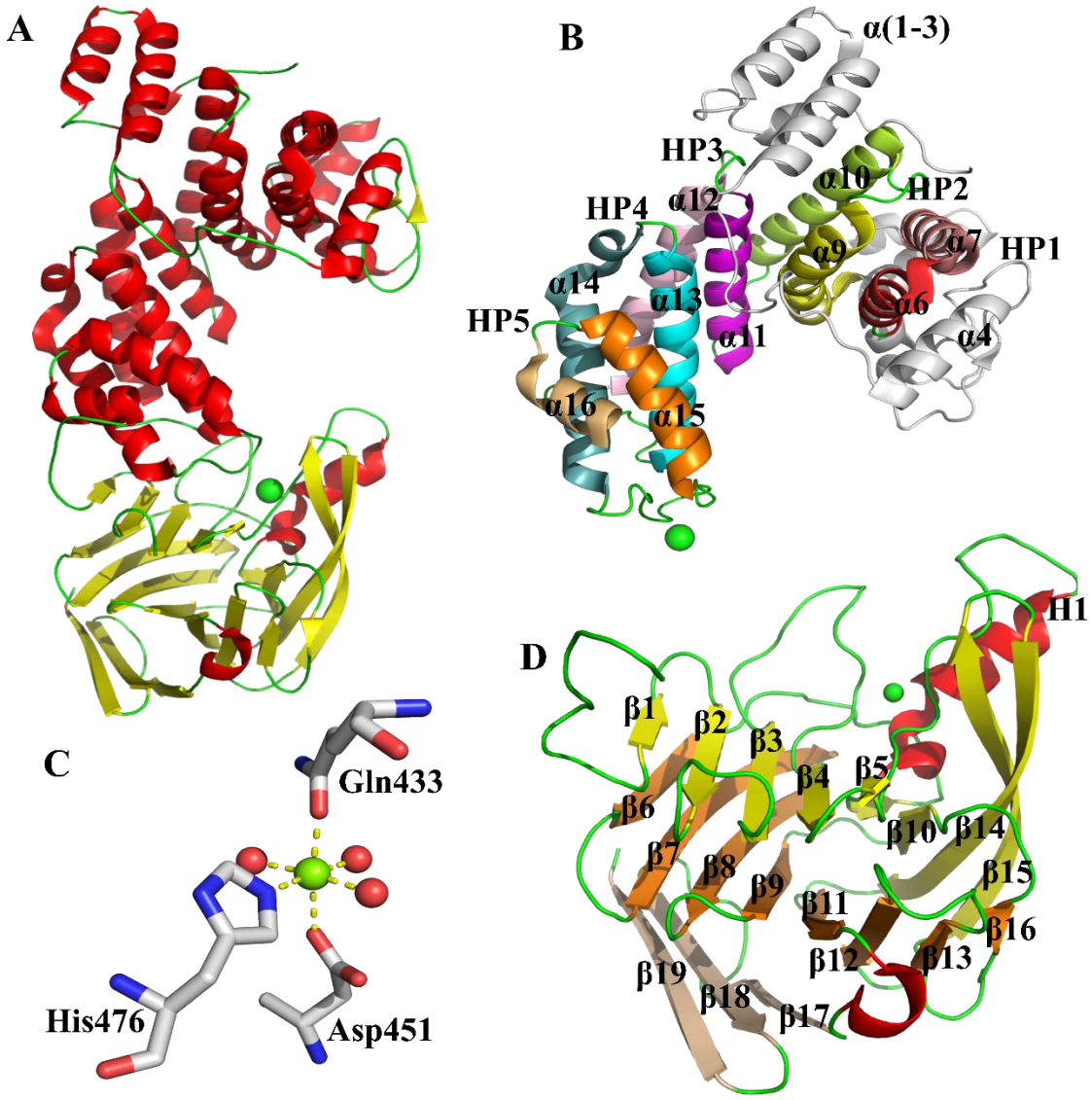

Figure 5:2 Structure description of HepIII.

A) Cartoon representation of HepIII showing α helices (red) in the N-terminal domain and β-sandwich (yellow) in the C-terminal domain; **B**) the N-terminal α-helical domain forming an incomplete $(\alpha/\alpha)_5$ toroid fold. Helices that are not part of the toroid are colored white, the hairpins (HP) are marked in different colors, the inner helix is a darker shade and the outer helix is a lighter shade. The metal ion (Mg^{2+}) is shown as green sphere; **C)** the Mg^{2+} ion and the liganding residues from the C-terminal domain are shown; **D)** the C-terminal β-sheet domain showing stacks of three β-sheets colored yellow, orange and wheat, respectively; the β-strands forming the three β-sheets are marked.

5.5 Active Site

Although HepIII and HepII share very little sequence identity and are classified into different families in the CAZy database, they both have the $(\alpha/\alpha)_n$ toroid+ β-sandwich fold. HepIII

was structurally similar to HepII (PDB code: 3E7J, 2FUT) from *Pedobacter heparinus*. HepII is capable of cleaving the C4-O1 bond next to either GlcA or IdoA epimers. The complex structure of HepII with tetrasaccharide substrate delineated the potential active site residues and the catalytic mechanism; the identified catalytic residues are His202, Tyr257, His406 and Glu205 (Shaya et al, 2006).

In HepII, the first step in the β-elimination mechanism (*i.e.* neutralization of the charge on the C6 carboxylic group) is done by Glu205 with the help of His406 and Arg261. For the cleavage of the GlcA epimer, Tyr257 plays both catalytic base and acid roles. For the cleavage of the IdoA epimer, His202 and Tyr257 play the catalytic base and acid roles, respectively. A structural comparison of the whole protein of HepIII and HepII does not align well. Structural superposition based on individual domains however, allowed us to draw some conclusions. The catalytically important residues Glu205, Tyr247, His406 of HepII have structural equivalents in HepIII namely Asn247, Tyr303 and His431, respectively. Aside from these similarities between HepII and HepIII active sites, there are also two notable differences. Namely, the absence of a His202 equivalent residue in HepIII and the presence of a second tyrosine sidechain, Tyr457 in HepIII facing the putative active site and opposite to Tyr301. Therefore, these residues N247, Y301, H431 and Y457 were selected as putative active site residues in HepIII **(Figure 5.3A)**.

5.6 Site-directed mutagenesis of active site and putative binding site residues

Four single mutants Y301F, H431A, Y457F and N247A, were made as described in section 4.2.1.5. The mutant proteins were purified in the same way as wildtype HepIII. The activity assay of the wildtype HepIII, methylated wildtype HepIII and the mutants on the heparan sulfate substrate was carried out as mentioned in section 4.2.4.5. This assay detected the activity of HepIII and showed that methylated HepIII is still able to depolymerize heparan sulfate albeit with reduced efficiency. Meanwhile, the Y457F mutation only moderately decreased the activity, suggesting that this residue is involved in substrate binding but not in catalysis. The other putative active site mutants (N247A, Y301 and H431) inactivated the enzyme, supporting their key role in catalysis. **(Figure 5.3B)**.

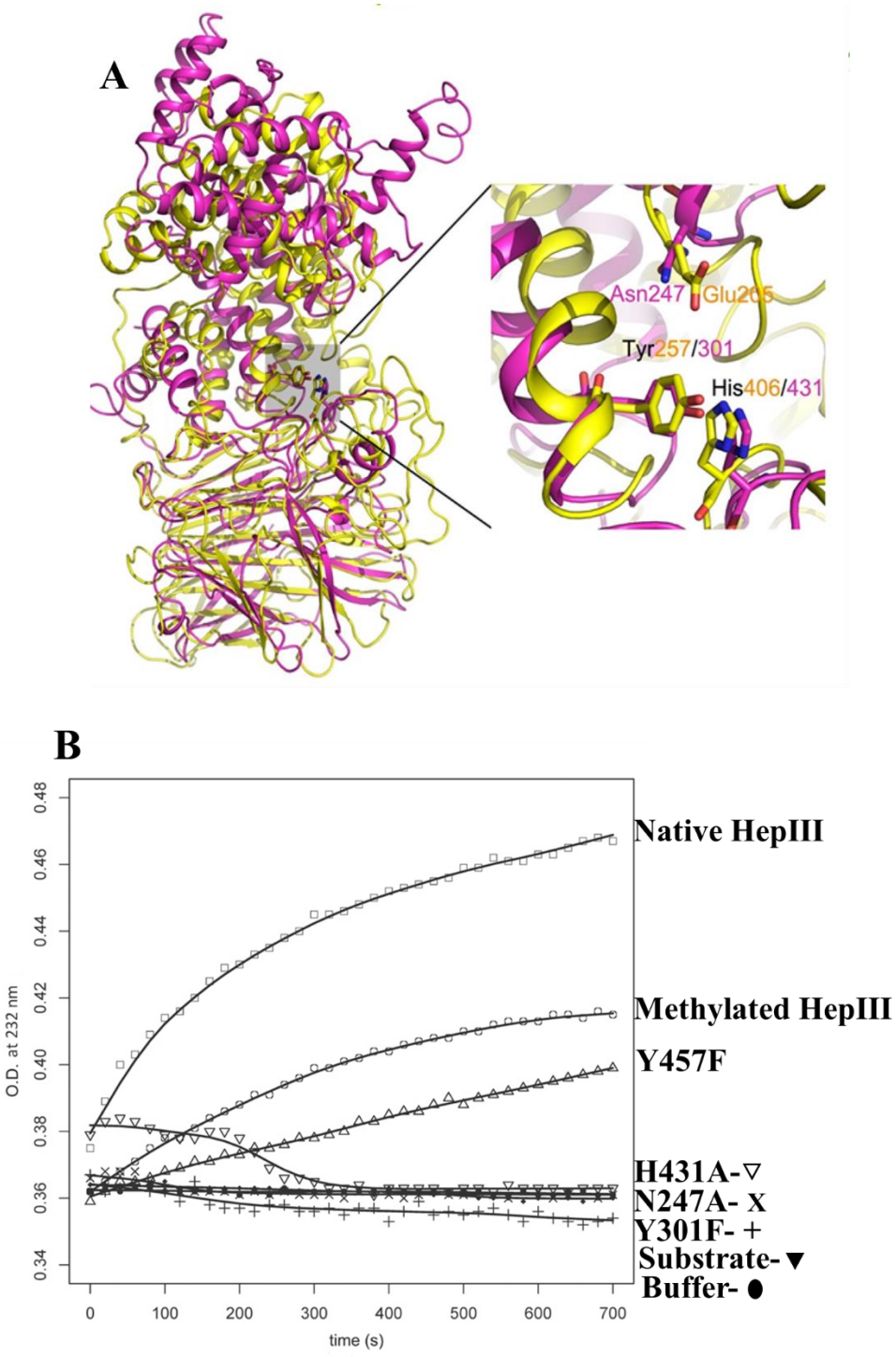

Figure 5:3 Putative active site residues and enzyme activity.

(A) Superposition of the HepIII (colored magenta) and HepII (yellow, PDB code: 2FUQ) structures based on the C-terminal domains. The insert shows the close-up of the active site Tyr301, His431 and Asn247 in stick representation. (**B)** Measurement of activity of the enzyme with heparan sulfate substrate. The absorbance at 232 nm was measured as a function of time.

5.7 Crystallization of HepIII mutant (Y301F)

The active site mutant (Y301F) of HepIII was purified in large scale for crystallization experiments. To obtain a protein ligand complex, efforts were made to co-crystallize the Y301F mutant with tetrasaccharide substrates (Section 4.2.4.6). Even after extensive screening for a crystallization condition, we were unable to obtain crystals of the protein–ligand complex. The methylated Y301F mutant on its own however, crystallized in multiple conditions, which happened to be slight variations on the condition where native HepIII crystallized: (1) 20%peg3350, 50 mM MgCl2, 0.1M Tris pH 8.5, 1 mM DTT with SB201 additive and in, (2) 25% peg3350, 0.2 M $MgCl_2$, 0.1M Tris pH 8.5, 5% glycerol, 1 mM DTT with SB195 additive. Tetrasaccharides substrates and long heparin digested by native HepIII were also soaked into the Y301F mutant crystals. For data collection, soaked crystals were transferred into a cryo solution containing 20% glycerol in mother liquor and flash cooled in liquid nitrogen. Diffraction data of the Y301F crystal was collected at the CLS, CMCF-ID beam line and the crystals diffracted to 3.6Å resolution. These crystals belong to space group $P2_1$ and contained one molecule in the asymmetric unit. However, the electron density map showed no density representing the tetrasaccharide substrates. In addition, soaking these crystals in the solution containing the substrate lowered the resolution of the diffraction pattern. Extensive attempts to obtain the structure of the HepIII (Y301F)-tetrasaccharide complex through either co-crystallization or soaking were not successful.

5.8 Structural comparison with homologs

While the author was working on crystallizing the HepIII-substrate complex**,** the structures of two other PL12 family enzymes were published; namely heparinase III (gene product of B*t*4662, PDB code: 4FNV) from *Bacteroides thetaiotaomicron* (Dong et al., 2012) and heparin sulfate lyase/heparinase III (HepC, PDB code: 4MMH) from *Pedobacter heparinus* (Hashimoto et al., 2014). Despite extensive trials, both groups were unable to capture the enzyme substrate/product/analog complex. Both structures were crystallized in the apo form and the putative active site residues were deduced by structural comparison with the HepII-substrate complex (PDB code: 3E7J). Comparison of the three HepIII structures from the PL12 family, as well as a comparison with HepII, allows us to rationalize the inability to capture the HepIII/substrate complex structure.

The homologs 4FNV and 4MMH share 35% and 40% sequence identity with HepIII, respectively. The structural comparison of 5JMD (our HepIII structure), 4FNV and 4MMH showed that the N- and C-terminal domains have similar arrangements, but the relative orientation of the domains differ between the three structures **(Figure 5.4A)**. The difference in the domain orientation is even observed in the structural comparison of two crystal forms of HepIII (5JMD and 5JMF-previously obtained crystal form in our lab) **(Figure 5.4B)**. The structural superposition of the N-terminal, α-helical domains of HepIII (residue 21-387) with 4FNV (residue 44-408) and 4MMH (residue 29-378) showed a root-mean-square deviation (RMSD) of 4.2 and 2.3 Å for 336 Cα atoms. In this structural superposition, the N-terminal helical cap and the first three α-hairpins superimpose well and the structures start to deviate for the fourth and fifth α-hairpins. This structural deviation in the last two α-hairpins defines the toroid fold as open (in the case of 4MMH) or closed (in the case of 4FNV).

Similarly, the structural superposition based on C-terminal, β-sheet domains of HepIII (390-666) with 4FNV (residue 424-702) and 4MMH (residue 392-659) showed an RMSD of 1.8 Å and 1.6 Å for 264 Cα atoms. Thus, the structural comparison of individual domains shows that the C-terminal domains are structurally better conserved than the N-terminal domains.

The superposition of entire proteins based only on their C-terminal β-sheet domains showed that the end of the second helix in the third α-hairpin, the following two antiparallel β-strands and the fourth (with the exception of the intra-hairpin loop) and fifth α-hairpins are rigidly attached to the C-terminal β-sheet domain **(Figure 5.4A)**. This connection between the N-terminal domain and the C-terminal domain may be solidified by the Mg^{2+}ion. Importantly, the active site residues are within this rigid region; Tyr301 is located at the beginning of the first helix in the fourth α-hairpin, and His431 is in the loop between third and fourth β-strands of the first β-sheet. **(Figure 5.5A)**. These structural comparisons give clues about the hinge region, which is located near Asn281 in HepIII, at the bottom of the second helix of the third α-hairpins (within the N-terminal domain).

The structural superposition of HepII (3E7J) and HepIII (5JMD), based on the C-terminal β-sheet domains, showed an RMSD of 1.58 Å for 171 Ca atoms (out of 247). The structural superposition based on the N-terminal α-helical domains showed good overlap of the first two α-helical hairpins (RMSD of 1.49 Å for 76 Ca), with the remaining three hairpins showing larger divergence. However, the fourth and fifth hairpins superimpose well with divergence of the first

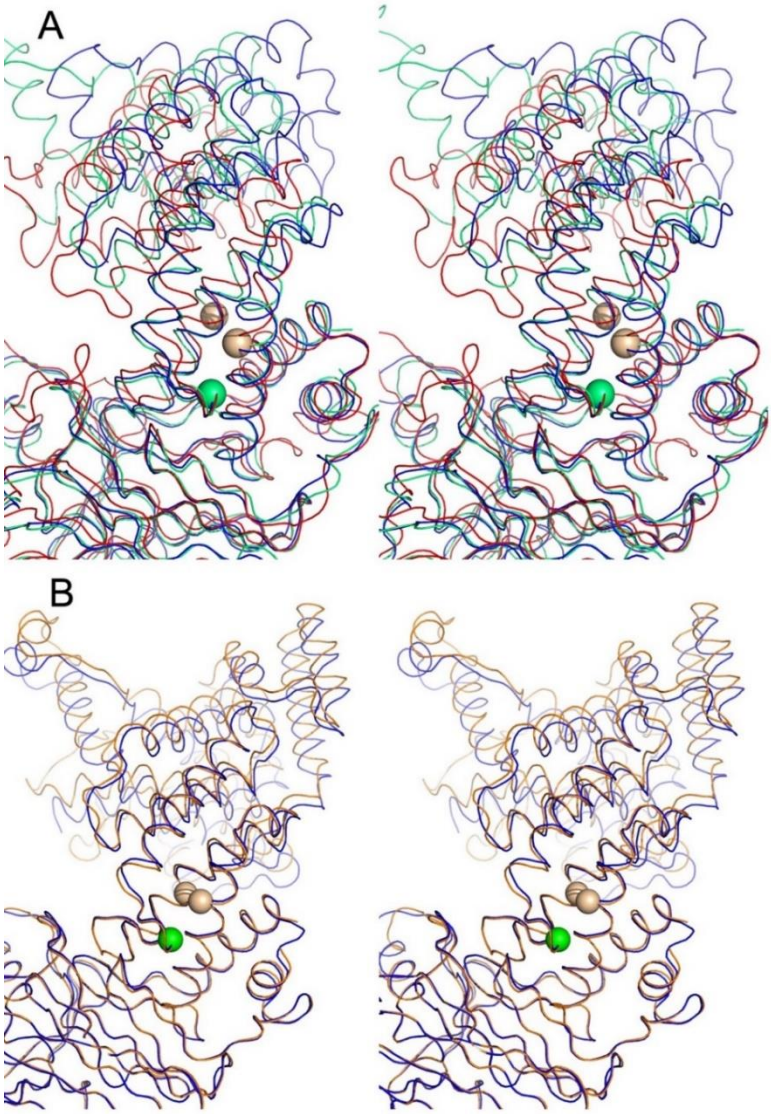

Figure 5:4 Superposition of various HepIII structures showing domain orientation. Superposition of the structures of heparinase III enzymes based on their C-terminal domains. The N-terminal domains were found in different orientations in these enzymes. (**A**) Comparison of 5JMD (green), 4FNV (red) and 4MMH (blue). In each case the chain preceding the first helical hairpin was removed for clarity. The spheres indicate the position of the hinge regions. The main hinge is located within the loop between the hairpins h3 and h4 (green sphere). Additionally, the inner loop within hairpin h4 with the tips of the helices (between the wheat-colored spheres) follows the movement of the N-terminal part of the protein; (**B**) Two crystal forms of HepIII (5JMD, 5JMF) superimposed based on their C-terminal domains. A small difference in the orientation of the N-terminal domains is easily visible. The hinge points are the same as when comparing different hepIII shown in **A**.

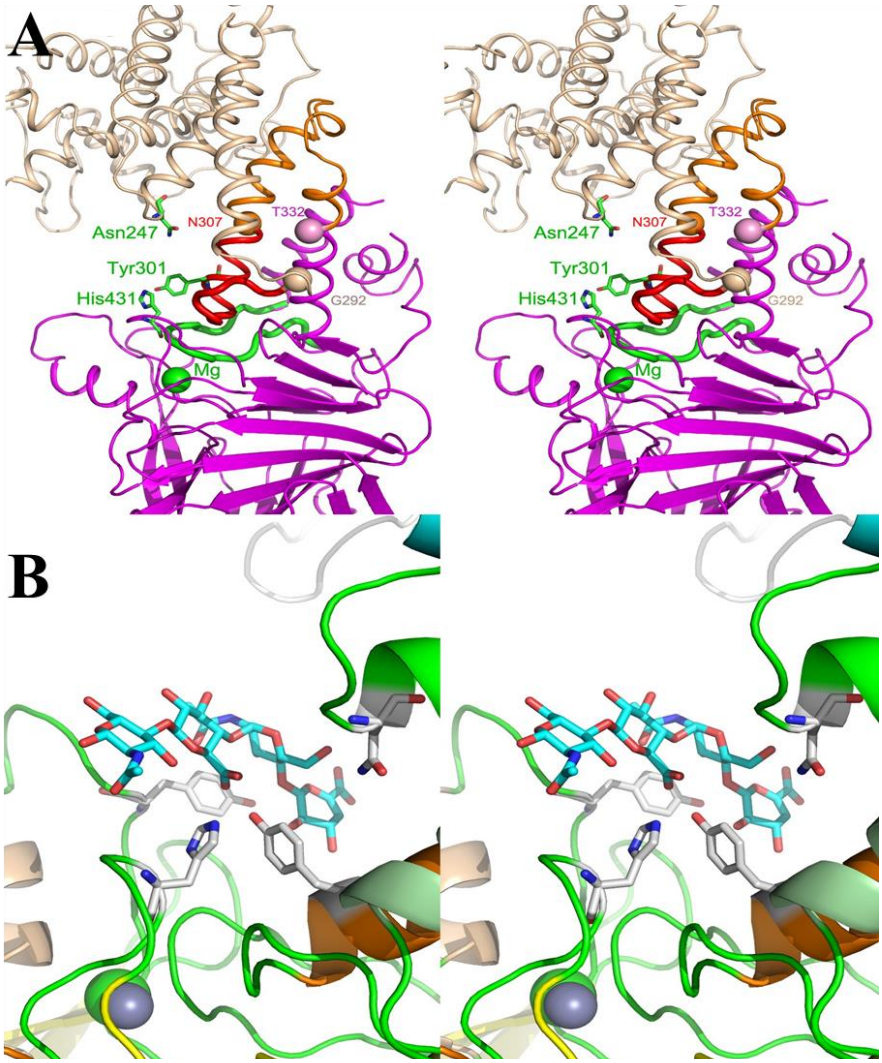

Figure 5:5 Location of putative active site residues and hinge point.
(**A**)Location of the active site residues (shown in stick mode, green) relative to the hinge points in HepIII (5JMD). The hinge points are marked as spheres. The N-terminal segment 23-291 (colored wheat) and residues 308-331 (orange) move together, while the end of α-helix and the following loop (aa 292-307, red) containing Tyr301 move together with the β-sheet C-terminal domain (magenta). The loop between hp4 and hp5 (aa 347-361, red) is tied in to the β-sheet domain by the metal ion (Mg); (**B**) Putative location of the tetrasaccharide substrate relative to the active site of HepIII (5JMD). The tetrasaccharide was extracted from HepII-tetrasaccharide (PDB code: 3E7J) complex structure upon superposition of the C-terminal domains.

three hairpins. This further supports the idea that differences in the arrangement of the five hairpins at the N-terminal domains is related to conformational changes around the third hairpin.

Also, in the structural superposition based on the C-terminal domains of HepII (3E7J) and HepIII (5JMD) the Tyr-His catalytic dyads, superimpose well. The location of the tetrasaccharide

in this superposition gives a good approximation of the expected location of the substrate in the active site of HepIII **(Figure 5.5B)**. However, it is evident from this superposition that the contacts of the substrate with the observed 'open' conformer of HepIII (5JMD) are rather limited, and there are no contacts with the other end of the open toroid. The same lack of contact of the substrate with the upper end of the N-terminal domain occurs for 4MMH that was also crystallized in an open conformation (Hashimoto et al., 2014). The 4FNV has been captured in a closed conformation and here the modeled tetrasaccharide would make contacts with the inner α-helix of the first α-hairpin (Dong et al., 2012). However, the lack of success in the crystallization of 4FNV with a substrate suggests that this conformation is also not compatible with substrate binding and catalysis.

5.8.1 Normal mode analysis

The three-dimensional structures of proteins from the PL12 family captured in the crystals display different conformations of their N-terminal α-helical domains and, based on our analysis described above, none of them represent the active conformation. This is different from enzymes belonging to several other PL families with the same fold, namely hyaluronan lyase, chondroitinase AC and heparinase II, which were crystallized in active conformations. To better understand the conformational dynamics of PL12 family enzymes, we employed molecular dynamics simulations. Since we were interested in large-scale movements rather than local dynamics, we opted for the NMA (Bahar et al, 2010). Starting from the crystal structures, we performed NMA for all three proteins. In each case, we found that the first two orthogonal normal modes describe the large-scale motions **(Figure 5.6 A-C**). The first mode represented a bending motion that brought the N- and C-terminal domains closer to each other, whereas the second mode represented a rotational motion that adjusted the orientation of the two domains. These motions occur within the N-terminal α-helical domain and the hinge region is located within the third α-hairpin. The hinge points in trajectories of all three HepIII enzymes are very closely mirroring the hinge points deduced from comparing the three crystal structures. We were further interested to see if the movement along the first two normal modes brings all three proteins to pass through a common conformation. This was indeed the case and the common conformation is shown in **(Figure 5.6D)**. This might be the active conformation. The position of the tetrasaccharide substrate

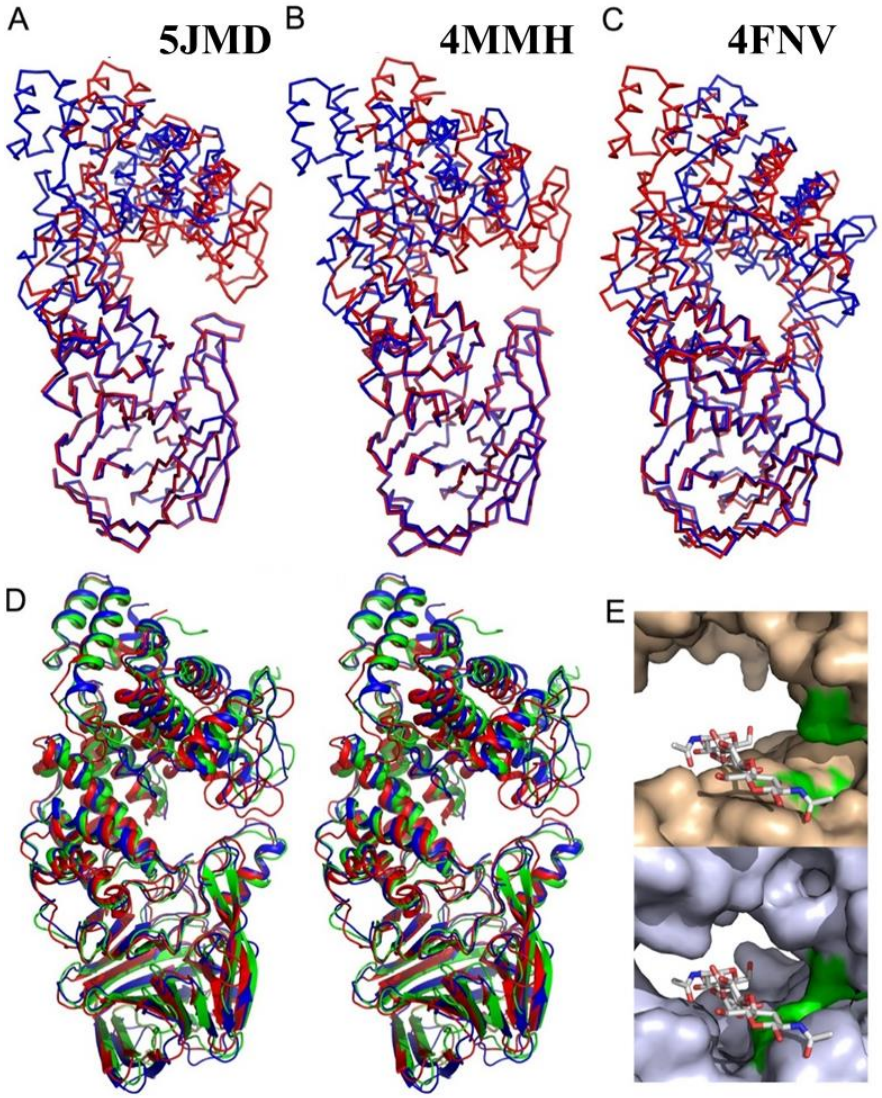

Figure 5:6 NMA and possible active conformation of HepIII.

Trajectory along the two normal modes of the NMA. All molecules are aligned using the C-terminal β-sheet domain only and shown in the same orientation. (A) Initial, open (blue, from the crystal structure) and the last, closed conformation (red) along the calculated trajectory for HepIII (5JMD); (B) the same for 4MMH, first - blue, last – red; (C) 4FNV, first - blue, last - red. Here the initial conformation is of the closed state and the final one is of the open state; (D) Stereo view of the conformations along the 51-frame trajectories of HepIII (5JMD) (blue), 4FNV (red) and 4MMH (green) that are closest to each other. They correspond to frame 25 for HepIII (5JMD), frame 20 for 4FNV and frame 51 for 4MMH and they approximate the active conformation of this family of heparinases; (E) The tetrasaccharide fitted into the observed conformation of HepIII (5JMD) (surface representation, top) and in the conformation common to all trajectories (bottom). The surface corresponding to the catalytically essential residues, Tyr301, His431 and Asn247, is colored green.

as transferred from HepII (3E7J), in the observed conformation of HepIII (5JMD) is shown in **(Figure. 5.6E, top)** and in the conformation common to all trajectories in **(Figure. 5.6E, bottom)**.

The fact that the all structures captured in the crystal are in an inactive conformation suggests that the lowest energy state of these lyases does not represent the conformation required for activity and that the shift to active conformation is driven by the energy derived from substrate binding. It might be that this conformational change is stimulated more easily by a longer heparan sulfate chain rather than by a tetra- or hexasaccharide that was used in the co-crystallization attempts.

Table 5-1 Data collection and refinement statistics of HepIII

Crystal form	Form 2
Space group	$P2_1$
a, b, c (Å), β(°)	66.5, 80.5, 83.3, 103.9
Wavelength (Å)	0.9796
Resolution (Å)	40.5-2.4
Observed hkl	155894
Unique hkl	33448
Completeness (%)	99.5 (98.5)
Redundancy	4.7
Rsym	0.130 (1.27)
I/(σI)	9.3 (1.6)
R_{work}	0.219
R_{free}	0.258
Wilson B ($Å^2$)	58.1
B-factor ($Å^2$, atoms)	
Protein	67.8 (5146)
Solvent	57.5 (122)
Ramachandran plot	
Favored (%)	94.3
Allowed (%)	5.1
Disallowed (%)	0.6
R.m.s. deviation	
Bonds (Å)	0.003
Angles (°)	0.6
PDB code	5JMD

6. Structure and catalytic mechanism of ulvan lyase enzymes

This section describes the structure and functional characterisation of the recently discovered and characterized ulvan lyases. The first identified and biochemically characterized ulvan lyases were NLR42 and NLR48 from *Nonlabens ulvanivorans*. They are currently classified within the PL28 family in the CAZy database. This was followed by identification and biochemical characterization of LOR107 (short ulvan lyase) & LOR61 (long ulvan lyase) from *Alteromonas* strain LOR. A blast search against the non-redundant protein database leads to the identification of LOR107 & LOR61 homologs in *Alteromonas* strain LTR [LTR_2195 (long ulvan lyase), LTR_2241 (short ulvan lyase)] and *Pseudoalteromonas* strain PLSV (PLSV_3925, PLSV_3875). These ulvan lyases are classified within the PL24 family in the CAZy database. Further bioinformatics analysis of the genome of *Pseudoalteromonas, Alteromonas and Nonlabens ulvanivorans* identified three novel ulvan lyases that share no sequence homology to the previously identified ulvan lyases. They are PLSV3936, LOR29 & NLR492. We have received all the aforementioned genes from our collaborators Drs. Ehud Banin (Bar-Ilan University, Israel) and William Helbert (CNRS, France). All these genes were cloned without signal peptides into an expression vector (pET28a) with a C terminal 6xHis tag. Since the structural fold and catalytic mechanism of these new families are unknown, we became interested in studying their structure, catalytic mechanism and evolutionary relationship to other lyases. This section is divided into three subsections as follows:

Subsection 6.1 titled "The structure and catalytic mechanism of LOR107 an ulvan lyase (PL24 family) from *Alteromonas* strain LOR". **Subsection 6.2** titled "The structure and catalytic mechanism of an ulvan lyase-PLSV3936 (PL25) from *Pseudoalteromonas* strain PLSV". **Subsection 6.3** titled "Structural characterization of NLR48 (Non-Classified PL category) from *Nonlabens ulvanivorans*".

6.1 The Structure and catalytic mechanism of ulvan lyase-LOR107 (PL24) from *Alteromonas* strain LOR

6.1.1 Expression and purification of LOR107

LOR107 was over expressed in BL21 (DE3)-RIPL cells and purified using TALON metal affinity resin as described in 2.2.5.1. The recombinant protein eluted with 40 mM imidazole was concentrated and loaded onto a S200 size exclusion column. The protein eluted as a single peak with an apparent molecular weight corresponding to a monomer **(Figure 6.1.1C)** was concentrated to 17 mg/ml and screened for crystallization. The results of affinity purification and SEC were evaluated by SDS-PAGE. **(Figure 6.1.1A, C)**.

6.1.2 Crystallization, Diffraction, Data Collection and Data Processing of LOR107

The wildtype LOR107 was screened for crystallization at room temperature as described in section 4.2.5.4. Initial crystals appeared in drops containing 25.5% PEG8K, 0.17 M ammonium sulfate, 0.085 M sodium cacodylate pH 6.5, and 15% glycerol as well solution. Optimization of this initial condition was carried out in hanging drops using 24-well grid screens. Crystals most suitable for diffraction were obtained after two weeks at room temperature in condition 23% PEG8K, 0.17M ammonium sulfate, 0.085M 4-Morpholineethanesulfonic acid (MES) pH 6.5 and 15% glycerol at a protein concentration of 20 mg/ml **(Figure 6.1.1B)**.

For data collection the crystals were cryoprotected by briefly immersing in the mother liquor and flash cooled in liquid nitrogen. Diffraction data were collected at the 08ID beamline (Grochulski et al., 2011) at the CLS and processed with the XDS program (Kabsch, 2010). These crystals diffracted to 1.9 Å resolution and belonged to space group $P2_12_12_1$ with unit cell parameters a=83.6, b=121.3, c=123.9 Å. Since there were no structures of homologs sharing high sequence identity with LOR107, we decided to use the single anamolous dispersion method to solve the structure. In preparation for this approach, selenomethionine (SeMet) incorporated LOR107 was purified and crystallized.

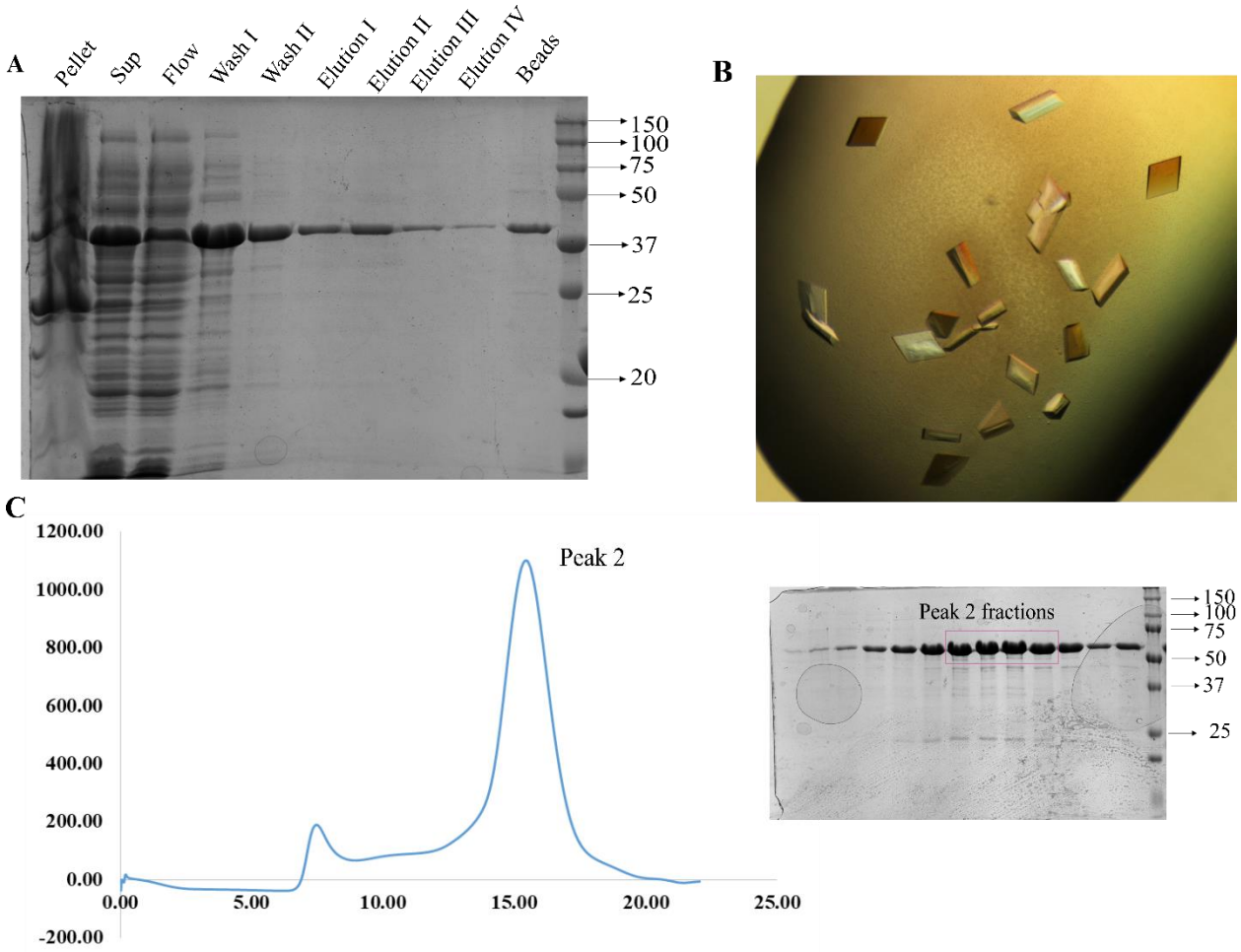

Figure 6.1:1 Purification and Crystallization of Wildtype LOR107.
(A) The coomassie stained SDS-PAGE analysis of LOR107 purified on cobalt affinity resin. Wash I and Wash II fractions corresponds to 20 mM Imidazole and 40 mM Imidazole in lysis buffer, respectively. Elution I, II are 250 mM and III, IV are 500 mM imidazole in lysis buffer, respectively. **(B)** Crystals of the LOR107 obtained from 23% PEG8K, 0.17 M ammonium sulfate, 0.085 M 4-Morpholineethanesulfonic acid (MES) pH 6.5 and 15% glycerol at room temperature. **(C)** SEC profile of LOR107 on GE S200 column. Peak2 elutes around ~15 ml represents the monomer behavior of LOR107. SDS-PAGE showing the purity of peak2 and the fractions used in the crystallization experiments (highlighted in pink box).

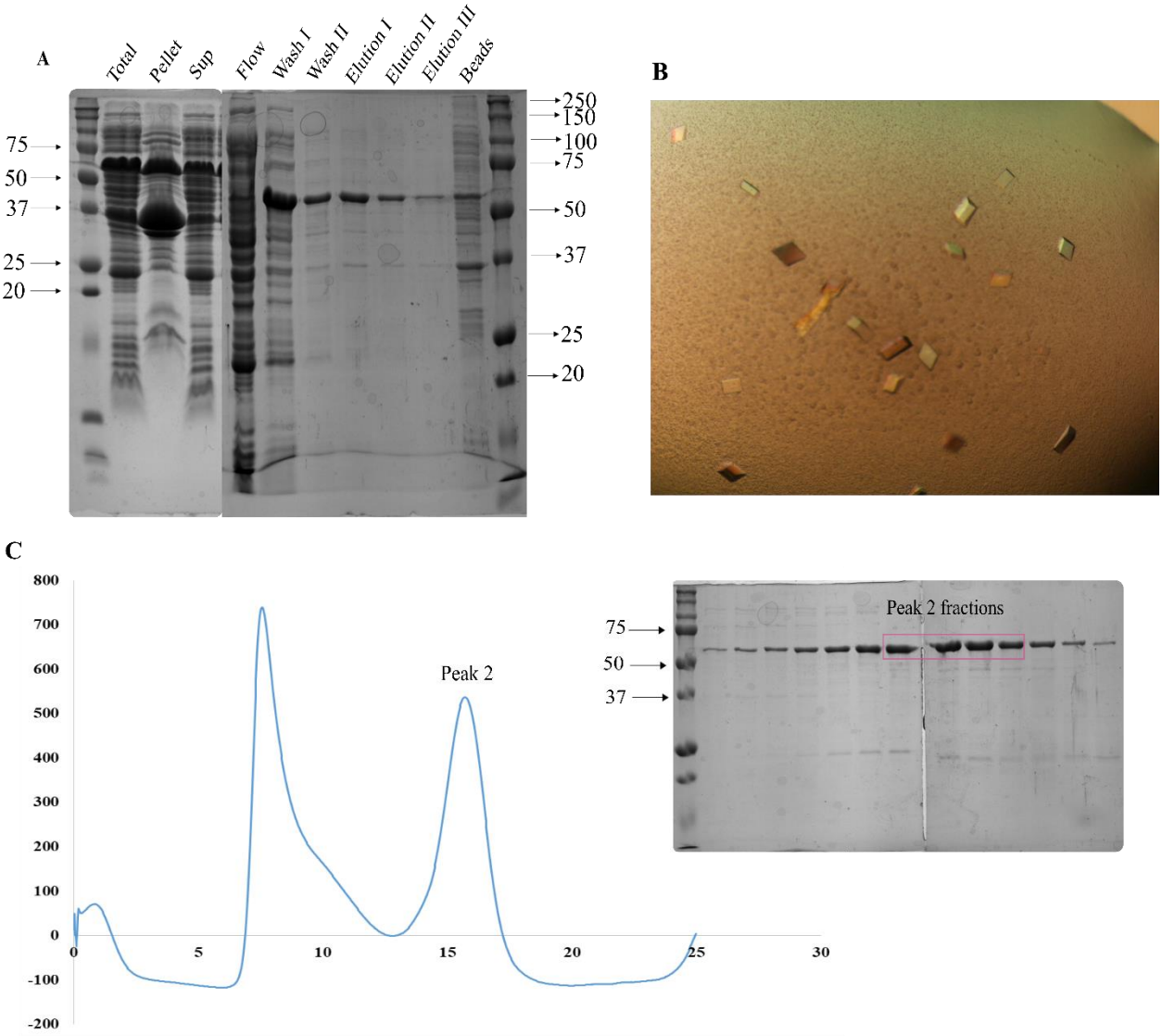

Figure 6.1:2 Purification and Crystallization of SeMet-LOR107.
(A) SDS-PAGE analysis on cobalt affinity purified SeMet-LOR107. Lysis buffer and the various fractions in the purification were all similar to wildtype purification. **(B)** Crystals of SeMet-LOR107 obtained from 25.5% PEG 8K, 0.17 M ammonium sulfate, 0.1 M MES pH 6.5 and 17% glycerol at room temperature. **(C)** SEC profile of SeMet-LOR107 on GE S200 column. Peak2 elutes around ~15 ml represents the monomeric behavior and SDS-PAGE showing the purity of peak2 fractions (fractions used in the crystallization experiments are highlighted in the pink box).

6.1.3 SeMet incorporation into LOR107, purification and crystallization

SeMet incorporated LOR107 (SeMet-LOR107) was produced by inhibiting the methionine biosynthesis pathway and purified as described in section 4.2.5.2 and 4.2.5.4. The purification and SEC profile of SeMet-LOR107 are shown in **(Figure 6.1.2A & C)**. Crystallization was carried in a 24-well grid screens using variations on the condition under which the native protein crystallized.

After optimization through microseeding the best crystals of SeMet-LOR107 were grown in 25.5% PEG 8K, 0.17 M ammonium sulfate, 0.1M MES pH 6.5 and 17% glycerol.

6.1.4 Diffraction, Data Collection and Data Processing of SeMet-LOR107

The SeMet-LOR107 crystals were briefly immersed in mother liquor for cryoprotection, followed by flash cooling in liquid nitrogen. Diffraction data were collected at the 08ID beamline at the CLS (Grochulski et al., 2011). Data was processed using HKL 3000 (Minor et al., 2006). The SeMet-LOR107 crystals diffracted to 2.5 Å resolution. These crystals belong to space group $P2_12_12_1$ with unit cell parameters a=84.2, b=121.9, c=124.3 Å. Data collection statistics are shown in Table 6-1.1.

6.1.5 Structure determination and refinement

The structure of LOR107 was solved with the automated structure solving pipeline CRANK2 (Skubák and Pannu, 2013) using data from the SeMet crystal. A total of 20 selenium sites were identified and the initial model contained 839 out of 1000 residues. This model was refined with the 1.9 Å resolution native crystal data. The structure was refined with the PHENIX refinement program (Adams et al., 2011) and manual rebuilding and solvent placement was conducted with the COOT program (Emsley and Cowtan, 2004). Two large peaks in a difference electron density map were observed in identical positions within each molecule. Based on the coordination and distances to the liganding atoms, these peaks were assigned as Ca^{2+} ions. The final model contains residues 40 to 520 in both chains, four Ca^{2+} ions, fourteen SO_4 ions, five glycerol molecules, one PEG (Polyethylene glycol) molecule and 753 water molecules. The refinement converged at R_{work}=0.175 and R_{free}=0.200. The stereochemistry of the model was validated with MolProbity (Chen et al., 2010). The coordinates were deposited in the PDB and the PDB code of wild type LOR107 is **6BYP**.

6.1.6 Crystal structure of LOR107

The protein crystallized with two molecules in the asymmetric unit. Both molecules appear to be identical and superimpose with a RMSD of 0.29 Å for all Cα atoms. The molecule adopts a seven-bladed β-propeller fold **(Figure 6.1.3A)**. Each propeller consists of four antiparallel β-strands.

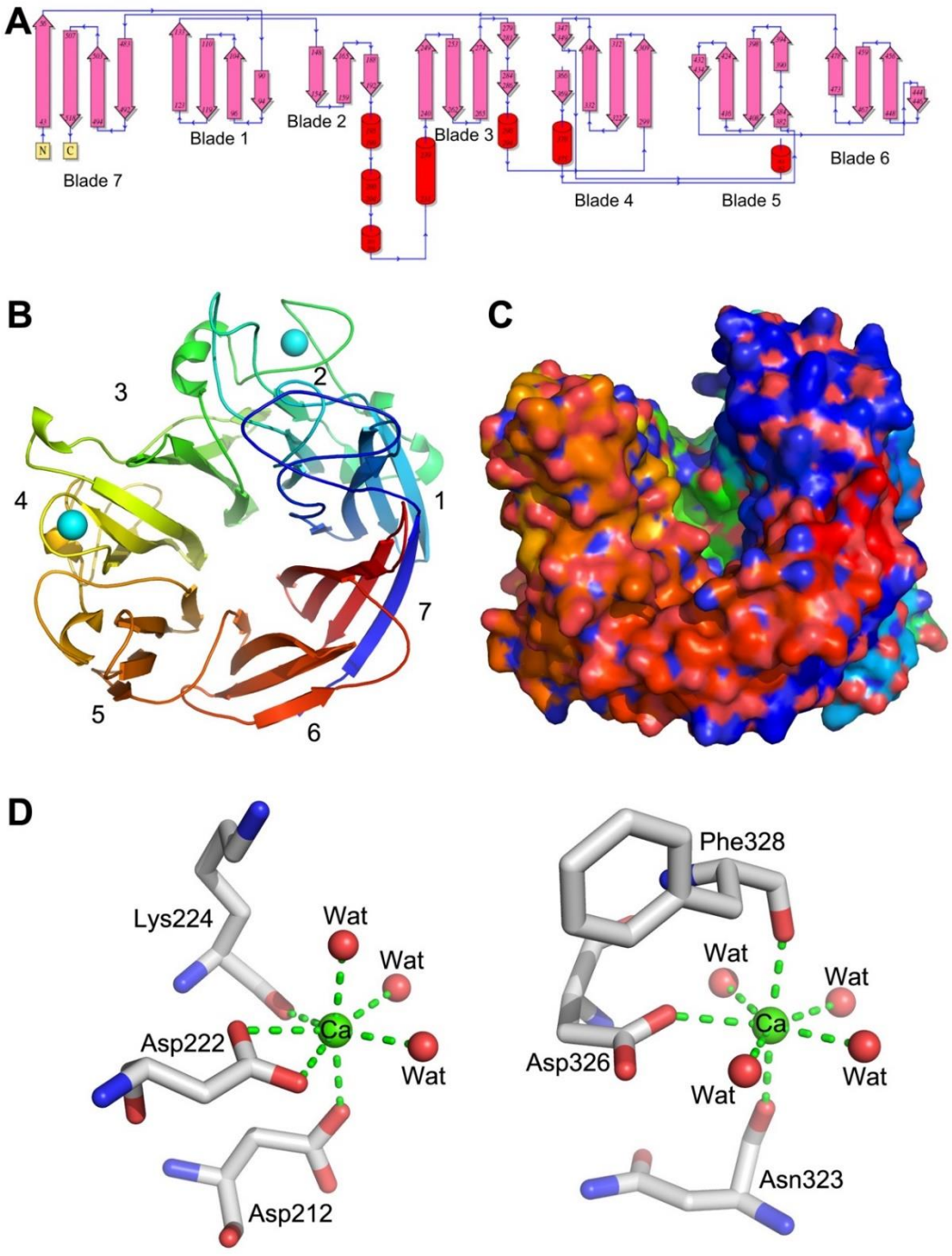

Figure 6.1:3 Structure description of LOR107.

The structure of LOR107. **(A)** Topology diagram of LOR107; **(B)** Cartoon representation of LOR107 painted in rainbow colors, from blue at the N-terminus to red at the C-terminus. Two Ca^{2} ions are shown as cyan spheres. **(C)** The surface representation showing a long canyon along its top surface with lower rims of the propeller at the ends of the canyon. The view is rotated by 70 degrees along horizontal axis relative to panel B. Color scheme as in panel B; **(D)** The Ca^{2+} ions in sites 1 (left) and 2 (right).

Blades 1, 2, 3, 4, 5 and 6 are formed by residue 90 - 133, 148 - 208, 240 - 286, 299 - 369, 382 - 434 and 444 - 478. The 7th blade is formed from three C-terminal β- strands, 483 - 518 and the 4th β-strand is provided by the N-terminal residues 43 - 56. The participation of both N- and C- terminal residues in the formation of a single blade in the β-propeller fold is a stabilizing phenomenon termed "molecular Velcro" (Munoz-Munoz et al., 2017). The fourth strand in blade 2 is distorted, forming fewer inter-chain hydrogen bonds. This strand is followed by a long loop that is stabilized by a metal ion. The loops joining β-strands extend on either side of the propeller. The loops on one side of the propeller are rather short (bottom of the propeller) while the loops on the opposite end are of varied length, with some extending high above the ends of the β-sheets (top of the propeller). Loops within blades 1, 2, 4, 5 and 7 are especially long, while those within blades 3 and 6 are much shorter. These long loops delineate a deep canyon on the top surface of the propeller, with high walls on two sides and a lower 'neck' located between blades 2 and 4 and over the short loops within blade 3 **(Figure 6.1.3C)**. The opposite end of the canyon also has a lower wall formed by blade 6. This canyon is lined with a cluster of highly conserved surface exposed residues, indicating that this surface defines the substrate binding site of LOR107.

The presence of metal ions in the structure was indicated by four very strong peaks in the difference electron density map. Two metal ions are present in each independent molecule at identical positions. Based on their coordination geometry these ions were identified as Ca^{2+}. The Ca^{2+} ions are located on the periphery of the propeller. Ca^{2+} in site 1 stabilizes the long loop leading from blade 2 to blade 3 and abutting the long loop between the distorted strand β4 of blade 2 and β1 of blade 3. This Ca^{2+} ion is surrounded by seven ligands in a pentagonal bipyramidal coordination. The equatorial ligands are provided by OD1 and OD2 of Asp222, a carbonyl oxygen of Lys224 and two water molecules, while the axial ligands are OD2 of Asp212 and another water molecule **(Figure 6.1.3D, left)**. The Ca^{2+} in site 2 is located in blade 4 and stabilizes a long loop between strands β2 and β3. The loop connecting blades 4 and 5 is leaning onto this loop. The coordination of this Ca^{2+} is also pentagonal bipyramidal, with equatorial ligands being OD2 of Asp326 and four water molecules, while the axial ligands are carbonyls of Asn323 and Phe328 **(Figure 6.1.3D, left)**.

6.1.7 Evolutionary analysis and active site architecture of LOR107

The evolutionarily conserved residues of LOR107 were identified using the program ConSurf (Landau et al., 2005). Two clusters of highly conserved residues were identified on the top of the propeller along the sides of the canyon, and contain polar, charged residues that could constitute the active site **(Figure 6.1.4A)**. Residues His146, His167, Thr242, Tyr243 and Arg259 belong to cluster I and are located on one side of the canyon at the inner edges of blades 2 and 3. Asn263, Tyr303, Arg320 and Tyr330 belong to cluster II and are on the opposite side of the canyon on the edge of blade 4 **(Figure 6.1.4B)**. These residues are highly conserved among PL24 sequences. The two clusters are separated by ~8-10 Å, defining the width of the canyon, which is sufficient to accommodate a linear ulvan polysaccharide. We have investigated the contribution of these amino acids to the enzyme activity of LOR107 by site-directed mutagenesis.

6.1.8 Active site mutants and Enzyme activity

The identified conserved residues were subjected to mutation studies. Single mutants H146A, H167A, T242A, Y243F, R259A, N263A, Y303F, R320A, and Y330F were made as described in section 4.2.5.3. The mutants were expressed and purified in the same way as the wildtype protein. The enzyme activity of the wildtype LOR107 and mutants on the ulvan substrate was measured as described in section 4.2.5.3. Mutation of H146A, H167A, R259A and R320A led to a complete loss of activity against the substrate. The Y243F mutant showed very low activity while the N263A and Y330F mutants showed reduced activity. Mutating Thr242 to an alanine led to a very small decrease in activity, while the Y303F mutant unexpectedly showed a small increase in activity relative to the wildtype LOR107. Together, these data suggested that His146, His167, Arg259 and Arg320 play essential roles during catalysis. Tyr243 is also important in catalysis, since the Y243F mutants showed only residual enzyme activity. The remaining investigated side chains most likely affect the organization of the active site or binding of the ulvan substrate **(Figure 6.1.4C)**

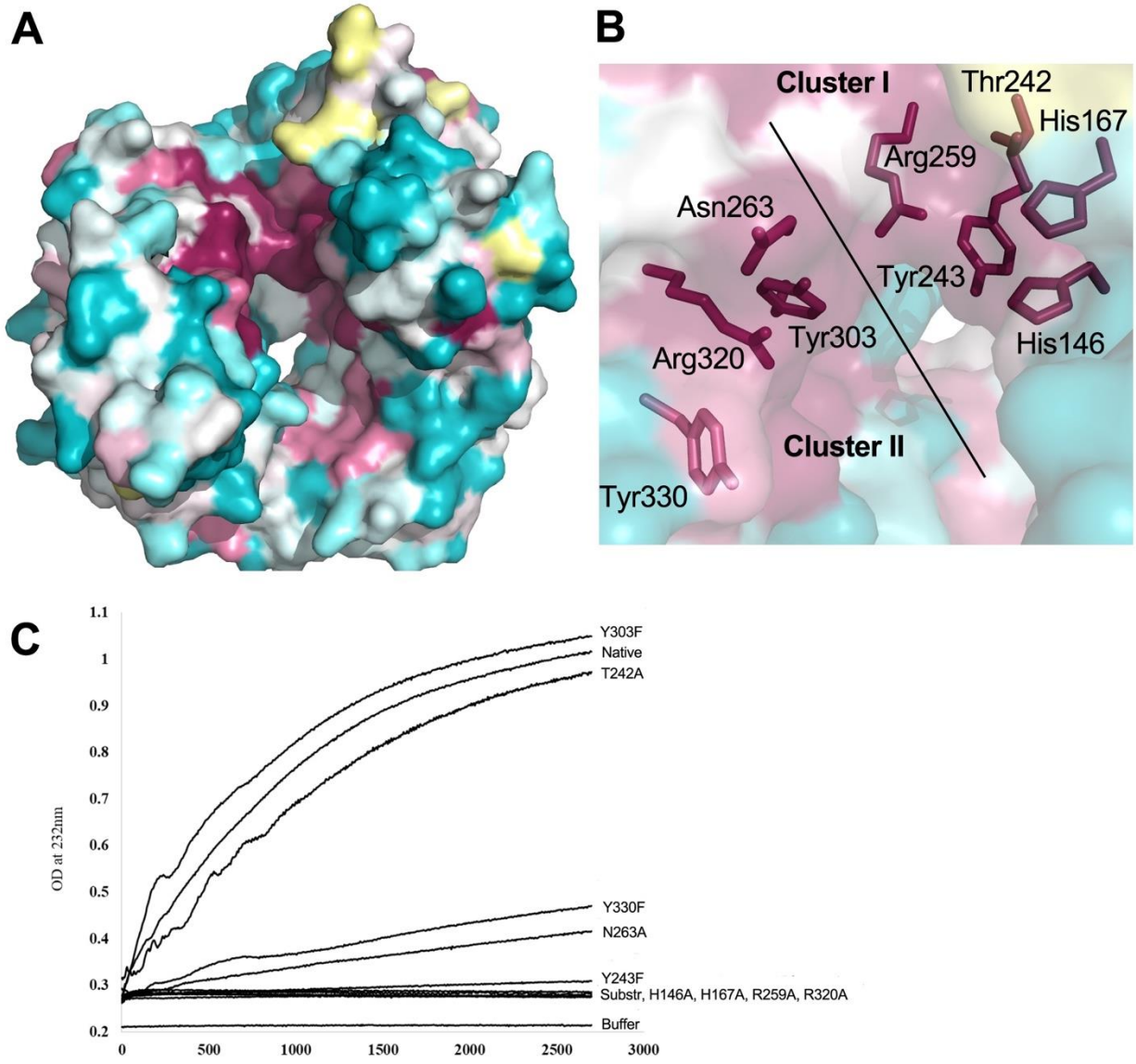

Figure 6.1:4 Sequence conservation and enzyme activity analysis of LOR107.
(A) The surface representation of the enzyme looking at the top of the propeller. The surface is colored by the level of sequence conservation among homologs. The burgundy color marks highly conserved and surface exposed residues. A cluster of highly conserved residues is visible near one end of the canyon; **(B)** The enlarged view of the conserved surface cluster. The surface is shown as semi-transparent and the conserved residues are shown in stick representation; **(C)** Activity of wildtype LOR107 and its mutants. Reaction conditions are 0.5mg/ml ulvan substrate and 1 µg of enzymes in 20 mM HEPES pH 7.5, 150 mM NaCl. Appearance of unsaturated sugars resulted from the lytic cleavage was measured at 232nm for 45 minutes.

.

6.1.9 Purification and crystallization of LOR107 mutants

To capture the substrate within the enzyme, several single mutants were made in which His146, His167, Tyr243, Arg259 and Arg320 were changed for other amino acids. His146 and H167 mutated to Asn or Leu did not express well and most of the protein was trapped in inclusion bodies. The successfully expressed mutants were Y243F, R259N and R320N. These mutants were purified in the same way as the native protein and concentrated to ~ 23 mg/ml. Crystallization of these proteins required optimization with the final condition being close to that of the native protein (25.5% PEG8000, 170 mM ammonium sulfate, 15 % glycerol, 0.1 M MES, pH 6.5).

6.1.10 Soaking experiments, diffraction and data collection of LOR107 mutant crystals

Crystals of LOR107 single mutants (R259N, Y243F and R320N) were soaked for ~5 hours in a solution containing two different tetrasaccharide substrates (section 4.2.5.7). Tetrasaccharide I (DP4-1) contained a mixture of ΔUA-Rha3S-GlcA-Rha3Sα/β and ΔUA-Rha3S-IdoA-Rha3Sα/β with some traces of hexasaccharide **(Figure 4.2)**. Tetrasaccharide II (DP4-2) was ΔUA-Rha3S-Xyl-Rha3Sα/β. After soaking in tetrasaccharide solution, the crystals were flash frozen in liquid nitrogen. Diffraction data of R259N, Y243F were collected at the 08B1 beamline (Fodje et al., 2014) and R320N were collected at the 08ID beamline (Grochulski et al., 2011) at the CLS. Data were processed with the XDS program (Kabsch, 2010). These crystals diffracted to 2.2 Å resolution and belong to space group $P2_12_12_1$ (were isomorphous to the native crystals). A strong positive difference electron density corresponding to the bound substrate was observed in R259N and R320N mutant crystals but not in the Y243F mutant soaked in DP4-1. No binding was detected for the crystal soaked in DP4-2. Data collection and refinement statistics are in Table 6.1.1. The coordinates of R259N and R320N complexes with the tetrasaccharide substrate were deposited in the PDB and the PDB codes are **6BYX** and **6BYT,** respectively.

6.1.11 Complex structure of LOR107 mutants with substrate

6.1.11.1 Structure of R259N mutant with tetrasaccharide substrate

The difference electron density map calculated for the R259N mutant showed an elongated blob of positive density located within the canyon on the top of LOR107 in proximity to the two clusters of conserved residues. We could fit the entire DP4-1 tetrasaccharide into this electron density **(Figure 6.1.5A, left panel)**. The GlcA could be fitted in the +1 position in a chair

conformation, while IdoA fitted only in a high energy boat conformation. In addition, since LOR107 shows specificity for cleavage next to GlcA, we have modelled and identified the uronic acid as GlcA in the +1 position. The substrate is located close to, and makes contacts with Arg320, Tyr303 and Asn263 (cluster II). Arg320 is in the center of this cluster and its sidechain makes hydrogen bonds to the +1 position $O3^{GlcA}$ and to -1 position $O5^{Rha3S}$. Arg320 is sandwiched between the two tyrosine sidechains conserved among PL24 sequences, Tyr303 and Tyr330, with the $C6^{Rha3S}$ methyl group abutting the face of the Tyr330 aromatic ring. Asn263, located next to the Arg320, forms H-bonds to $NH1^{Arg320}$ and contacts the +1 position $C2^{GlcA}$ proton through the aliphatic part of its side chain. In addition, Tyr303 is hydrogen-bonded to the +1 position $O6B^{GlcA}$ of the acidic group and the $O6A^{GlcA}$ is ~5 Å from Tyr243. Finally, His384 hydrogen-bonds to the -2 position $O6A^{\Delta UA}$ and Val386 faces the hydrophobic side of the -1 position Rha3S. Residues His146 and Tyr243 of cluster I lie nearest (~4-5Å) to the +1 position $O6A^{GlcA}$ of the tetrasaccharide substrate. Additionally, Arg177 H-bonds to the +2 position $O2^{Rha3S}$ **(Figure 6.1.5A, right panel)**. The mutated sidechain, Asn259 is far from the tetrasaccharide, but when this structure is superimposed on the native LOR107, the sidechain of
Arg259 extends toward the carboxylic group of GlcA at the +1 position and could form a salt bridge with it. The tetrasaccharide in the complex described above is somewhat distant from His146, His167 and Tyr243, which are important for catalysis, suggesting that this binding site is not catalytically competent. We therefore investigated the R320N mutant where the key cluster II contact residue has been changed.

6.1.11.2 Structure of the R320N mutant with tetrasaccharide substrate

In the R320N substrate complex, the substrate is positioned close to cluster I residues **(Figure 6.1.5B, left panel**). Unexpectedly, the Arg259 sidechain assumes two conformations in chain A. The minor conformer (occupancy 0.4) corresponds to that of the wild-type enzyme and forms a salt bridge with the acidic group of GlcA at the +1 position, while the major conformer (occupancy 0.6) points away from GlcA and is directed toward the backbone of Gly262-Asn263. His146 points toward the +1 position $H5^{GlcA}$ and the $NE2^{His146}$ is only ~2 Å away from this proton (~3Å from $C5^{GlcA}$). In addition, this NE2 nitrogen forms an H-bond with the bridging O4 between +1 and -1 sugars. Tyr243 also points toward C5 of GlcA and the bridging O4 but the distances are ~4 Å, too long for a direct participation in proton transfer. The SO3 of Rha3S at the -1 position

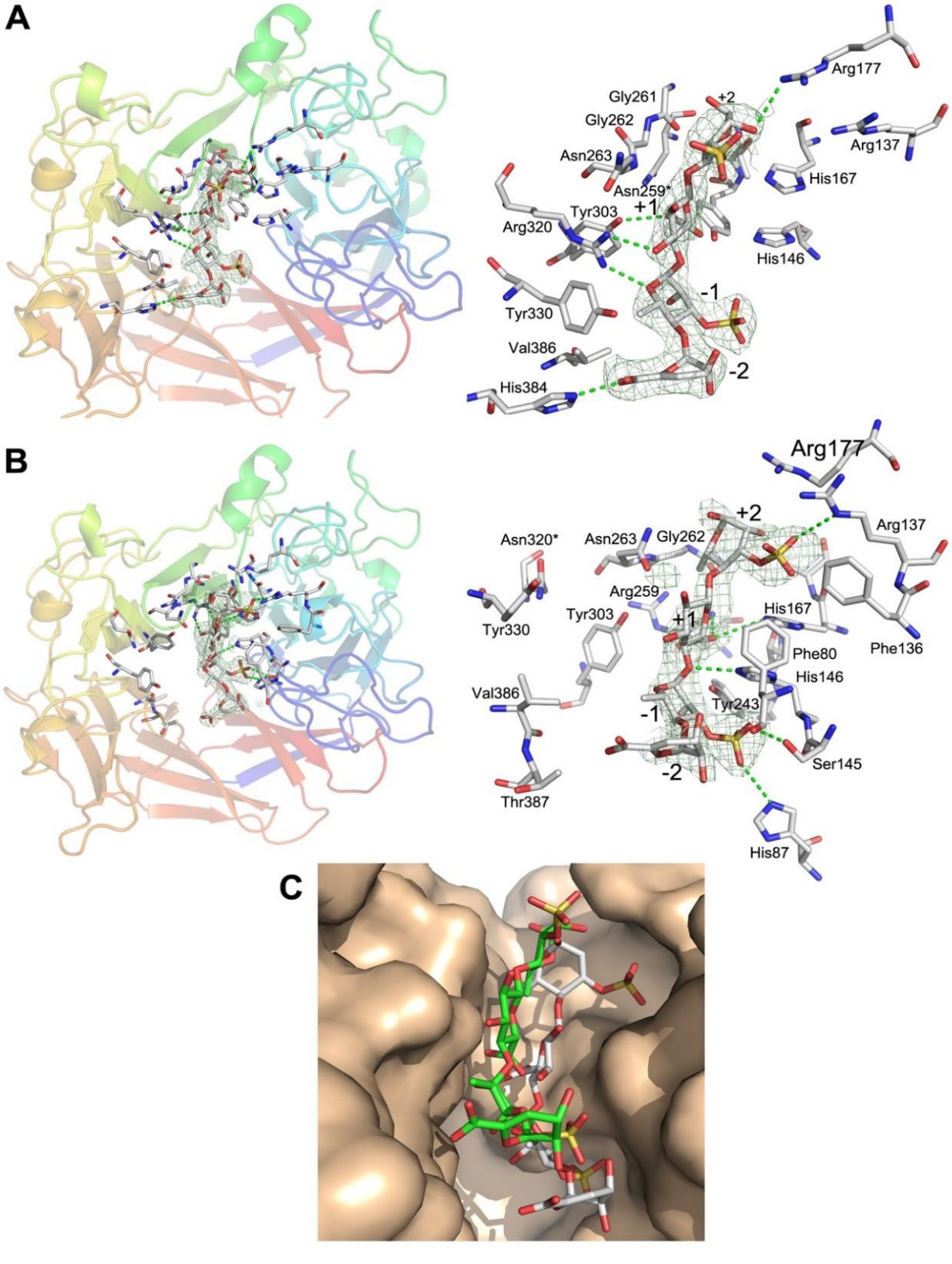

Figure 6.1:5 Structures of R259N and R320N mutants with tetrasaccharide substrate.

(A) Location of the tetrasaccharide within the canyon on the top of the propeller in R259N (left panel). The omit electron density map of the tetrasaccharide in R259N mutant contoured at 2.8σ with the model of the tetrasaccharide. The contacts with the protein are marked with dashed lines. Right panel-a close-up of the tetrasaccharide with the surrounding residues shown in stick representation; **(B)** Location of the tetrasaccharide within the canyon on the top of the propeller in R320N (left panel). The omit electron density map of the tetrasaccharide in the R259N mutant contoured at 2.8σ. The contacts with the protein are marked with dashed lines: Asn263 changes the conformation clamping the substrate. Right panel-a close-up of the tetrasaccharide with the surrounding residues shown in stick representation: **(C)** Superposition of the tetrasaccharides from R259N and R320N complexes and their location in the substrate binding canyon. Tetrasaccharide from R259N complex is colored green and R320N complex is colored gray.

fills a hydrophilic pocket made by Ser145, Tyr243 and His146 side chains. Similarly, the SO3 of +2 position Rha3S fits into a depression formed by Arg137, Arg177, His167 and the Phe137 side chains. Moreover, Asn263 clamps the tetrasaccharide from the opposite side by H-bonding to the oxygen atom that bridges +1 and +2 sugars. The unsaturated sugar in the -2 position is located near the main chain of Phe80 and Gly81 and makes only van der Waals interactions with the protein **(Figure 6.1.5B, right panel)**.

The R259N and R320N mutations had minimal effect on the overall structure of the protein. However, there is a significant difference in the way the tetrasaccharide substrate binds to these mutants **(Figure 6.1.5C)**. In the R259N complex, the tetrasaccharide substrate binds cluster II sidechains, interacting largely with Arg320, and is located at the upper part of the canyon, distant from the residues affecting catalysis. Nearby Asn263 is H-bonded to Arg320 and directed toward the top of the canyon. In the R320N complex the tetrasaccharide binds deeper in the canyon and approaches the cluster I sidechains. The Asn263 changes conformation and extends toward the other wall of the canyon making it narrower and pinning the substrate to the active site.

6.1.12 Putative Catalytic Mechanism

The importance of Arg320 in cluster II for bond cleavage, combined with the structure of the LOR107 (R259N)-tetrasaccharide complex, suggests that the initial binding of the substrate in the canyon is coordinated by Arg320. This binding site is near the top of the canyon and is distant from the residues essential for catalysis. We postulate that, upon initial binding the Arg259 side chain extends toward the acidic group of the +1 position GlcA and this promotes movement of the substrate deeper into the canyon toward the cluster I residues. The associated movement of the Asn263 sidechain narrows the canyon and locks the substrate in the active site. Indeed, the Asn263

side chain plays a role in the catalytic mechanism since its mutation to an alanine has a detrimental effect on activity **(Figure 6.1.3C)**.

In the R259N mutant, the Arg259 side chain is absent and the substrate remains in the initial binding site. However, when the Arg320 side chain is missing but Arg259 is present (i.e. the R320N mutant), the substrate goes directly to the final destination site near the bottom of the canyon. This binding site is largely preformed when compared with the native LOR107 structure. In the absence of Arg320 the tetrasaccharide substrate binds in what appears to be a catalytically competent or nearly competent conformation. We postulate that in the wild type LOR107, the movement of substrate deeper into the canyon pulls on Arg320 and causes a change in the conformation of Asn263, narrowing of the canyon and forcing the substrate into the catalytically competent conformation.

The catalytic mechanism of polysaccharide lyase family enzymes follow three steps: 1) neutralization of the negative charge on carboxylic acid of the uronic acid sugar. This reduces the pKa of the C5 proton; 2) abstraction of the C5 proton by a base; 3) donation of proton to the leaving group by an acid. A comparison of the wild type and R320N structures indicates that the role of the neutralizer is performed by Arg259 with the help of His167 **(Figure 6.1.6A)**. We suggest that the presence of His167 is essential for correctly orienting the acidic group on GlcA for the interaction with Arg259. The role of proton abstraction is likely performed by His146, which is properly oriented by stacking with His167 and interaction with Tyr243. The $ND2^{His146}$ atom becomes close to both $H5^{GlcA}$ of +1 GlcA and to the bridging O4 (to Rha3S), suggesting that His146 might shuttle the proton directly to the bridging O4 **(Figure 6.1.6B)**. The fact that the Y243F mutant displays residual activity indicates that Tyr243 plays a supporting role but is not directly involved in proton transfer. Therefore, LOR107 possesses an active site consistent with the His/Tyr mechanism, one of the two mechanisms identified so far in polysaccharide lyases, but applies a modification to this mechanism, in which the tyrosine plays a supporting rather than a direct role in proton shuffling **(Figure 6.1.6B)**.

The R320N mutant is inactive and this reflects the fact that either the substrate position observed in the R320N mutant or the arrangement of catalytic site residues are not fully compatible with catalysis. In this mutant, the Arg259 side chain is predominantly directed away from the substrate and toward the backbone of Gly262-Asn263, likely due to a small rearrangement of a

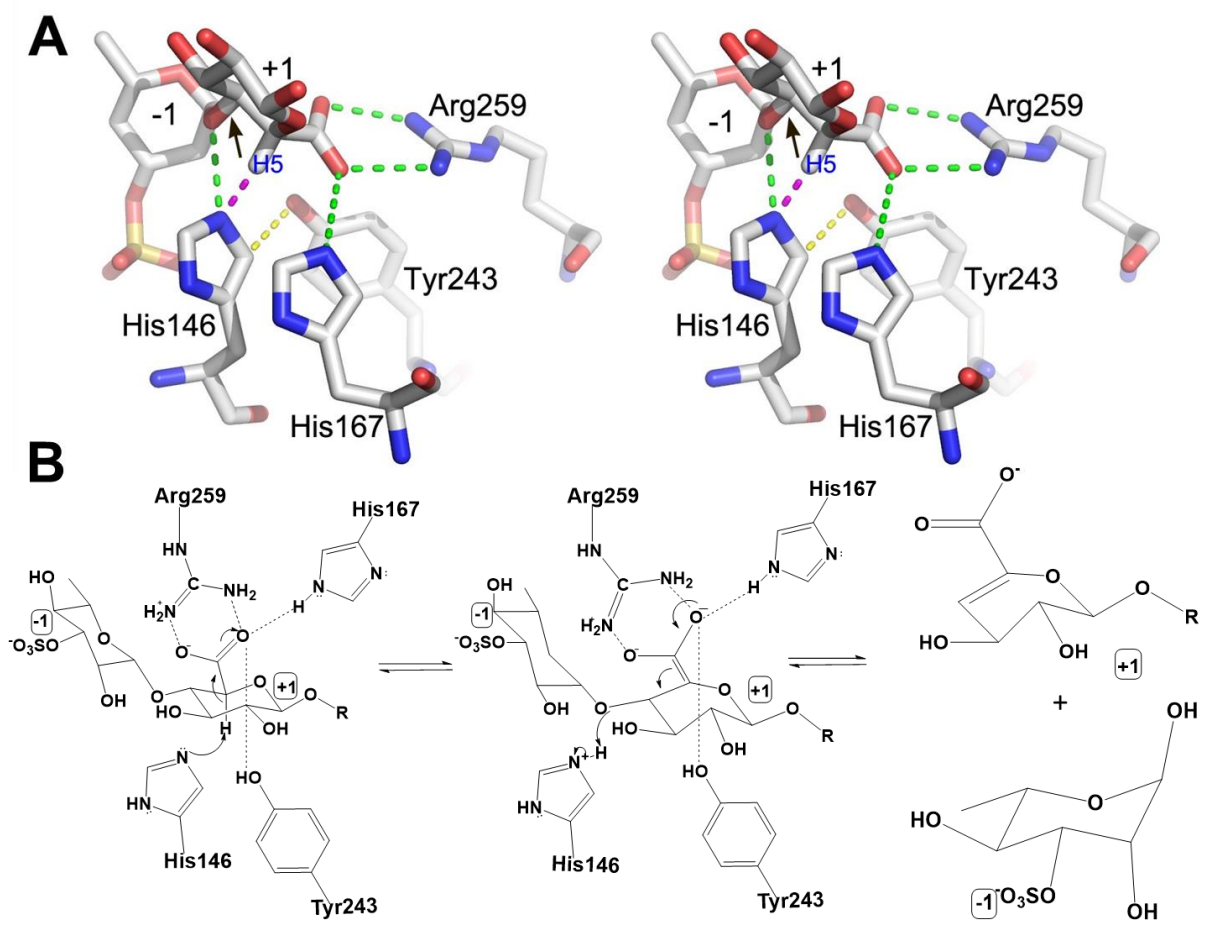

Figure 6.1:6 Proposed catalytic mechanism of LOR107.
Proposed catalytic mechanism. **(A**) A stereo view of the close-up of the central disaccharide showing the surrounding sidechains. The hydrogen bonds are shown as green dashed lines. The magenta dashed line indicates the 2.0 Å distance from NE2His146 to H5 proton of GlcA. The yellow dashed line indicates a short distance of ~3 Å between His146 and Tyr243. Tyr243 could form a H-bond to ND2 when His146 flips the sidechain by 180^{o}; **(B)** The proposed mechanism. Arg259 acts as a neutralizer of the charge of GlcA acidic group. His146 is the most likely candidate to play the role of acid and base by abstracting the proton from C5 of GlcA and donating it to the bridging O4. Tyr243 plays a supporting role in properly orienting the imidazole ring of His146. His167 contributes by stacking with His146 and also H-bonding to the acidic group of GlcA for its proper alignment in the active site.

neighboring segment near Asn263. Therefore, Arg259 cannot neutralize the acidic group of +1 GlcA, which in turn, prevents the C5 proton abstraction.

To summarize, we solved the crystal structure of LOR107, the first atomic structure of the PL24 family. LOR107 adopts the seven bladed β-propeller fold. The complex structures of two different active site mutants (R259N, R320N) with the tetrasaccharide allowed us to postulate the

possible catalytic mechanism. The catalytic cycle starts with a two-step binding event, first substrate interacts near the top of the canyon coordinated by Arg320, followed by sliding of the substrate into the canyon toward the active site residues. This is followed by the neutralization of negative charge on the acidic group of GlcA by Arg259 with the help of His167. Meanwhile, His146 plays the role of general base and/or general acid.

Table 6.1-1. Data collection and refinement statistics of LOR107

Crystal form	Native LOR_107	SeMet LOR_107	Complex (R259)	Complex (R320)
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
a, b, c (Å)	83.6, 121.3, 123.8	84.2, 121.9, 124.3	83.05, 121.0, 127.2	84.2, 120.7, 127.1
Wavelength (Å)	0.9795	0.9794	0.9787	0.9796
Resolution (Å)	49.07-1.90 (1.95-1.90)	46.26-2.5 (2.57-2.50)	48.89-2.21 (2.26-2.21)	49.05-2.20 (2.26-2.20)
Observed hkl	1497238	661588	954175	344075
# Unique hkl	99600	85354	64618	66028
Completeness (%)	99.9 (99.2)	99.8 (98.1)	99.0 (87.3)	99.9 (99.8)
Redundancy	15 (14.9)	7.8	14.8 (13.5)	5.2 (5.2)
Rsym	0.107 (1.94)	0.135 (1.35)	0.105 (1.32)	0.101 (1.25)
CC1/2	99.9 (65.2)	99.8 (64.7)	100.0 (91.0)	99.9 (53.6)
I/(σI)	16.56 (1.72)	13.27 (2.12)	23.8 (2.63)	12.93 (1.58)
R_{work}	0.175		0.20	0.20
R_{free}	0.20		0.25	0.25
Wilson B ($Å^2$)	33.7		36.05	39.3
Ramachandran plot				
Favored (%)	96.66		94.87	95.41
Allowed (%)	3.34		4.81	4.49
Clash score	4.25		8.84	7.82
R.m.s. deviation				
Bonds (Å)	0.006		0.007	0.008
Angles (°)	0.802		0.981	1.02
PDB code	6BYP		6BYX	6BYT

6.2 The Structure and Catalytic Mechanism of PLSV3936 from *Pseudoalteromonas*

Genome mining of ulvanolytic bacteria revealed the organization of co-regulated PULs. Functional characterization of the enzymes encoded by PULs lead to the identification of another ulvan lyase family unrelated to the PL24 enzymes (Foran et al., 2017). These new ulvan lyases were found in *Alteromonas LOR* (LOR29), *Nonlabens ulvanivorans* (NLR492) and *Pseudoalteromonas PLSV* (PLSV3936). They have been added to the PL25 family in the CAZY database. These three enzymes share ~ 70% sequence similarity with one another, but less than 14% with LOR107. We attempted to crystallize these proteins to gain insight into the folds and catalytic mechanisms of ulvan lyases with unrelated sequences.

6.2.1 Purification of PL25 family (PLSV3936, LOR29, NLR492)

Constructs of PLSV3936, LOR29 and NLR492 lacking the signal peptide were overexpressed in BL21(DE3)-RIPL cell lines. The proteins were purified as described in section 4.2.5.1. The fractions were eluted with 40mM Imidazole in lysis buffer, concentrated and analyzed by SEC. The results of affinity purification of PLSV3936 are shown in **(Figure 6.2.1A).** PLSV3936, LOR29 and NLR492 behave as monomers on SEC and the fractions corresponding to this monomeric peak **(Figure 6.2.1C)** were concentrated to ~30mg/ml and used for crystallization screening.

6.2.2 Crystallization trials of PLSV3936, LOR29 and NLR492

All the three enzymes were screened for crystallization by the sitting drop method as described in section 4.2.5.4. The protein NLR492 did not crystallize and LOR29 formed thin plate or rod-shaped crystals in several conditions but these crystals could not be improved. However, PLSV3936 also crystallized in several conditions with Peg 3350 as a precipitant and in the presence of various salts (0.2 M ammonium acetate, 0.2 M potassium iodide, 0.2 M sodium formate, and 0.2 M ammonium formate). Optimization of PLSV3936 crystals by the hanging drop vapor diffusion method improved the quality of these crystals **(Figure 6.2.1B)**. For data collection, crystals grown in 20% PEG3500, 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 5.5 were briefly soaked in cryo protectant (70% Mother liquor and 30% Ethylene glycol) and flash cooled in liquid nitrogen. Diffraction data was collected at the 08B1 beamline at the CLS (Grochulski et al, 2011) and processed with XDS (Kabsch, 2010).

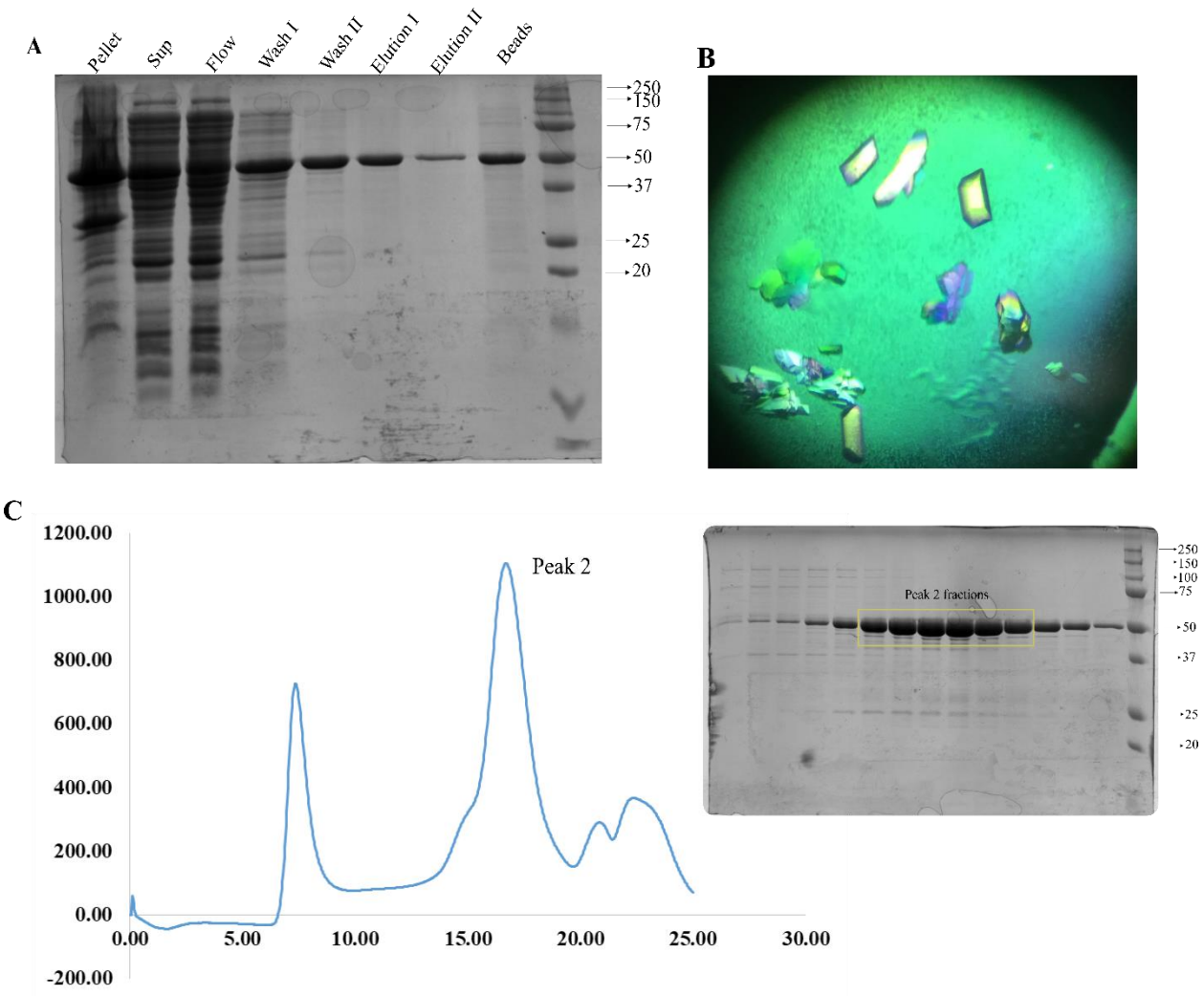

Figure 6.2:1 Purification and Crystallization of Wildtype PLSV3936.
(**A)** The coomassie stained SDS-PAGE analysis of PLSV3936 purified on cobalt affinity resin. Lysis buffer used in the purification is 20 mM Tris pH8, 200 mM NaCl, 5% glycerol and 5 mM Imidazole. Wash I and Wash II fractions corresponds to 20 mM Imidazole and 40 mM Imidazole in lysis buffer respectively. Elution I, II are 250 mM and 500 mM imidazole in lysis buffer respectively. **(B)** Crystals of His-PLSV3936 obtained from 20% PEG3500, 0.2 M Ammonium acetate, 0.1 M Bis-Tris pH 5.5 at room temperature. **(C)** Size exclusion chromatogram of PLSV3936 on S200 column. Peak2 elutes around ~15.8 ml represents the monomeric behavior. SDS-PAGE showing the purity of peak2 and the fractions used in the crystallization experiments (highlighted in the yellow box).

These crystals diffracted to 1.45 Å resolution and belonged to space group *C*2 with unit cell parameters a=174.3, b=72.1, c=109.5 Å and β=122.5º. There are two molecules in the asymmetric unit. Since PLSV3936 shares ~ 14% sequence identity with LOR107, the structure could not be solved by molecular replacement. SeMet incorporated PLSV3936 (SeMet-

PLSV3936) was therefore prepared to facilitate the structure determination by anomalous phasing methods.

6.2.3 Purification and crystallization of SeMet- PLSV3936

SeMet- PLSV3936 was produced by inhibiting the methionine biosynthesis pathway (Walden, 2010) as described in section 4.2.5.5. Purification and SEC analysis of SeMet-PLSV3936 are shown in **(Figure 6.2.2A, C).** The SeMet-PLSV3936 was concentrated to 30 mg/ml and used for crystallization. After optimization around the wildtype protein crystallization condition, the best crystals of SeMet-PLSV3936 were grown in 20% PEG3500, 0.2M ammonium acetate, 0.1M MES pH 6.5 (**Figure 6.2.2B)**. These crystals were briefly transferred to mother liquor containing 30% Ethylene glycol and flash cooled in liquid nitrogen. Diffraction data was collected at the CLS (08B1 beamline) (Fodje et al., 2014). The crystals diffracted to 2.05 Å resolution and the data was processed using XDS (Kabsch, 2010). The crystals belong to space group *C*2 with unit cell parameters a=174.2, b=71.7, c=109.4 Å and β=122.6^{o}. Data collection statistics are summarized in Table 6-2.1.

6.2.4 Structure determination and refinement

The structure of PLSV3936 was solved with the AUTOSOL program (Terwilliger et al., 2009)(available in the PHENIX software) using data from the SeMet incorporated crystal. There are two molecules in the asymmetric unit and each contains 10 methionine residues. A total of 23 anomalous scatterers were found and the initial model contained 787 out of 978 residues. This model was refined against the native dataset (1.45 Å resolution) using PHENIX software (Adams et al., 2011) with manual inspection and rebuilding using COOT (Emsley et al., 2010). A strong peak was observed in the mFo-DFc difference electron density map at identical positions, in each molecule. Based on the coordination and distances to the liganding atoms these peaks were identified to be Zn^{2+} ions. The fluorescence scan at the beamline confirmed the presence of Zn^{2+} ion in the crystal and these ions were also identified as anomalous scatterers in the single-wavelength anomalous diffraction (SAD) experiment. The final model contains residues 47 to 486 in chain A, 48 to 487 in chain B, two Zn^{2+} ions, two SO_4 ions, one PO_4 ion, one glycerol, nine ethylene glycol molecules and 1129 water molecules. The final refined structure has R_{work}=0.134 and R_{free}=0.143. The stereochemistry of the model was checked with PHENIX software and the

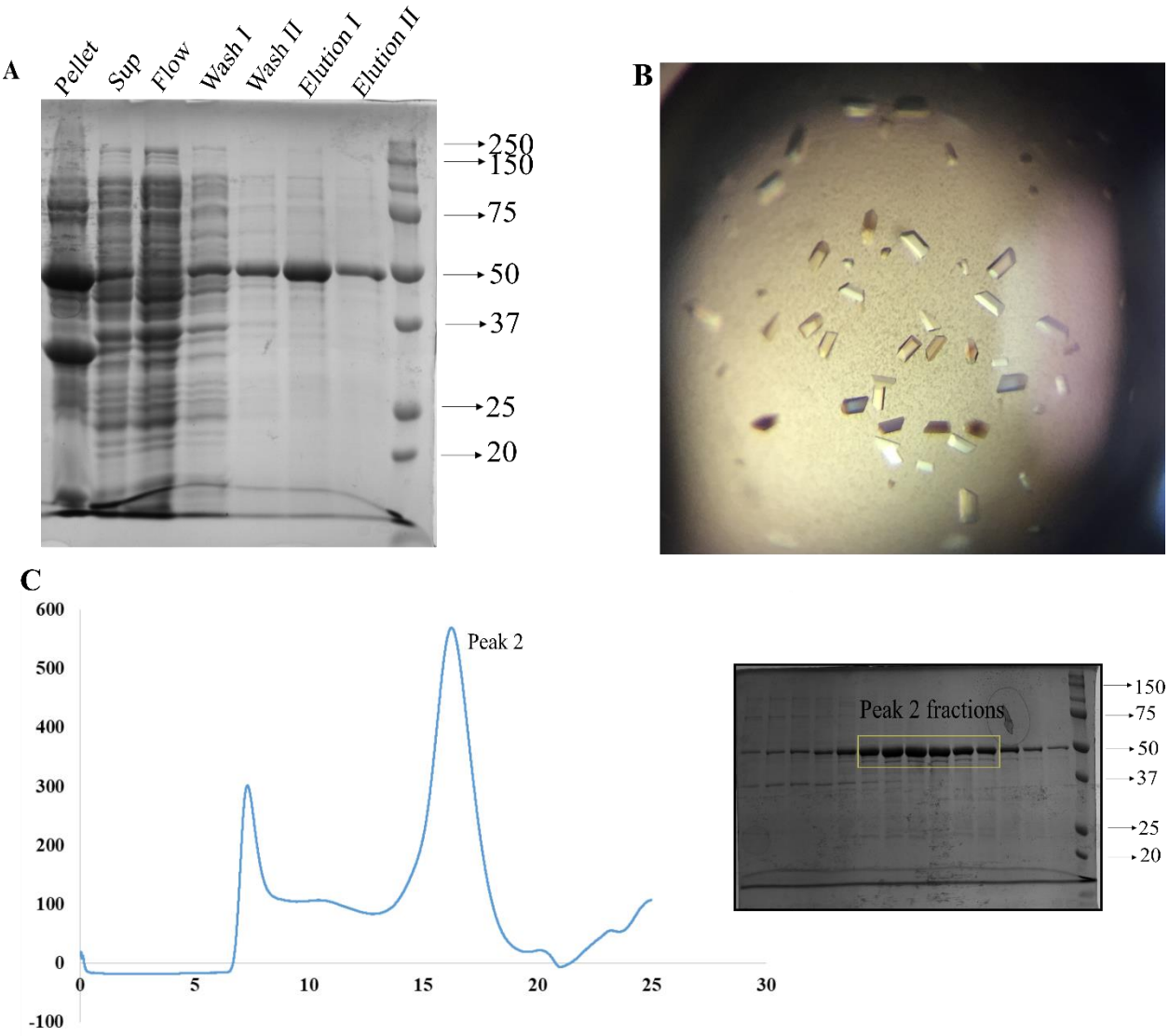

Figure 6.2:2 Purification and Crystallization of SeMet-PLSV3936.
(A) SDS-PAGE analysis on cobalt affinity purified SeMet-PLSV3936. Lysis buffer and the various fractions in the purification were all similar to wildtype purification. **(B)** Crystals of SeMet-PLSV3936 obtained from 20% PEG3500, 0.2 M Ammonium acetate, 0.1 M MES pH 6.5 at room temperature. **(C)** SEC profile of SeMet-PLSV3936 on S200 column. Peak2 elutes around ~15.8 ml represents the monomeric behavior and SDS-PAGE showing the purity of peak2 (fractions used in the crystallization experiments are highlighted in the yellow box).

refinement statistics are shown in Table 6-2.1. The coordinates and structure factors have been deposited in the Protein Data Bank with accession number **5UAM** (for the native structure).

6.2.5 Crystal structure of PLSV3936

The protein crystallized with two molecules in the asymmetric unit and they superimpose with a RMSD of 0.21 Å for all Cα atoms. The molecule adopts the fold of a seven-bladed β-

propeller **(Figure 6.2.3A)**. Each propeller consists of four antiparallel β-strands. Blades 1, 2, 3, 4, 5 and 6 are formed by residues 69 - 106, 125 - 180, 185 - 233, 240 - 321, 327 -378 and 385 - 434. The 7th blade is formed from three C-terminal β-strands, 435 - 474, with the 4th β-strand provided by the N-terminal residues 51 - 59. The loops joining the β-strands are short on one side of the cylinder (bottom), while the connecting loops on the opposite side (top) are of varied length and significantly longer. In particular, the 4th and 5th blades have long extensions on top of the propeller and the connection between the 1st and 2nd blade is also long. As a consequence, one end of the cylinder is flat, with only a small depression in the center while the other end has a deep and elongated crevice. The walls of the crevice on two sides are made of long extensions and are high. Conversely, the wall of the crevice over the 3rd blade is lower, forming a shallow neck as an entrance to the crevice **(Figure 6.2.3B)**. A sulfate molecule and several metal ions are bound within this crevice, making it the 'business end' of the lyase where the substrate binds. A Zn^{2+} ion is bound between the 2nd and 3rd blade in both molecules in the same place, liganded by the sidechains of His208, His 264, His278 and Cys266 in a tetrahedral coordination. The distances to the histidine NE2 or ND1 atoms 2.05 Å and to SG of cysteine 2.27 Å (soft restraints were used in the PHENIX refinement) **(Figure 6.2.3C)**.

6.2.6 Identification of Catalytic site and site directed mutagenesis:

Although LOR107 and PLSV3936 share only ~14% sequence identity, they adopt the same seven bladed β-propeller fold, with each blade having four antiparallel β strands. The putative catalytic site of PLSV3936 was identified based on a structural comparison with LOR107 and the nature of the residues in the binding pocket **(Figure 6.2.4A)**. We were interested in obtaining inactive mutants that we could use to capture the enzyme-substrate complex. To that end, the following mutants were made: H143N, H123N, R204N, H264N and Y188F. All these mutants were purified in the same way as the native enzyme. The activity was measured for the native enzyme and each mutant. All mutants except H143N, showed no measurable activity against the ulvan substrate **(Figure 6.2.4B)**. This indicates the involvement of these residues in either catalysis or substrate binding.

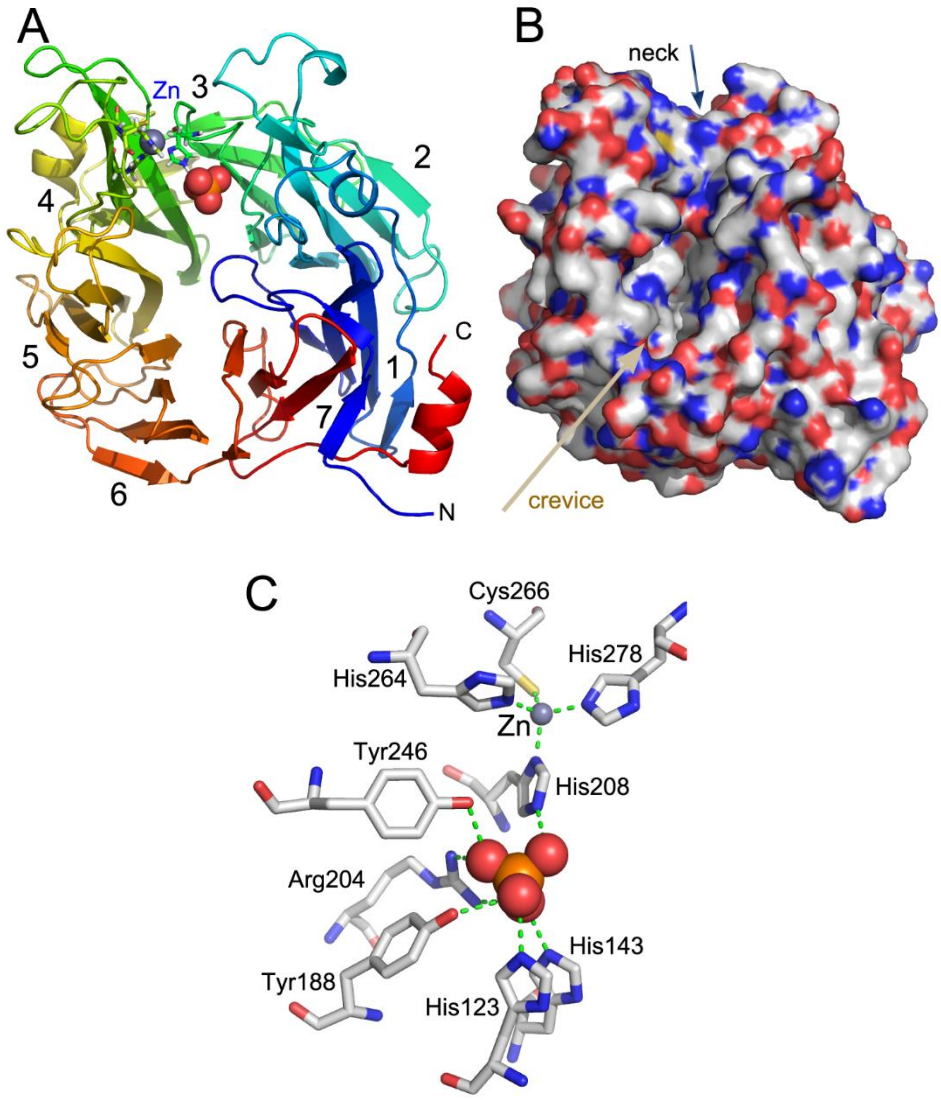

Figure 6.2:3 Structure features of PLSV3936.

(A) Overall structure of PLSV3936 in cartoon representation and colored in rainbow pattern from blue at N terminus to red at C terminus. The seven blades of β-propeller are named from 1-7. **(B)** Surface representation in same orientation as Figure **A**, pointing out the neck and crevice. **(C)** Zn $^{2+}$ and phosphate ions environment and the residues making contact with are shown in stick representation. (ThirumalaiSelvi Ulaganathan et al., 2017)

6.2.7 Purification and crystallization of PLSV mutants (H123N, H264N, Y188F)

The three mutants H123N, Y188F, H264N were purified in the same way as the native protein. Purified H123N, H264N and Y188F mutant proteins were concentrated to 29, 29 and 17

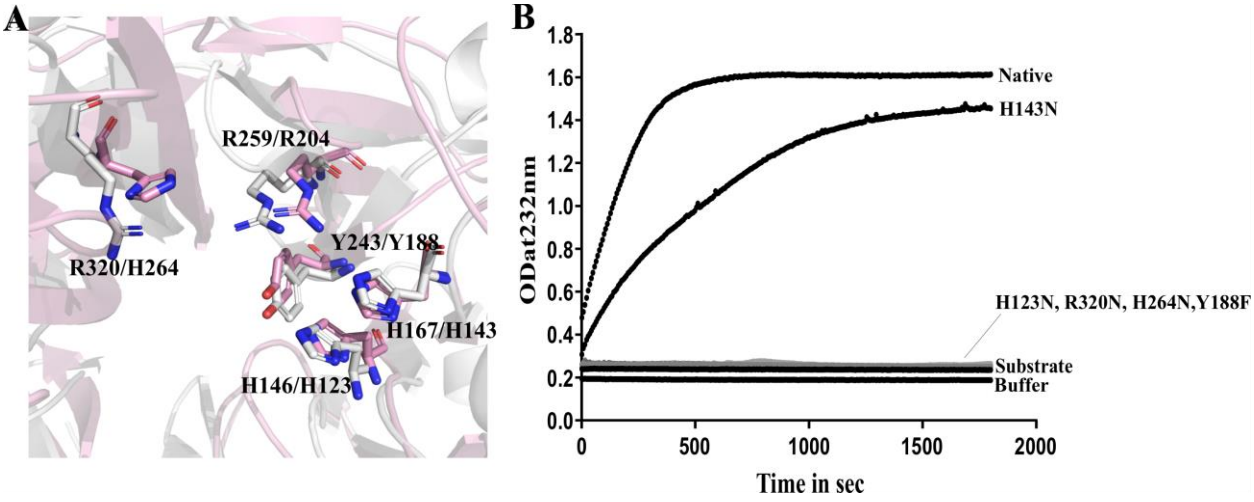

Figure 6.2:4 Identification of putative active site in PLSV3936.
(A) Active site residues of LOR107 (colored white) and their corresponding structure equivalent in PLSV3936 (colored pink), numbering of amino acid is LOR107/PLSV3936. **(B)** Enzyme activity plot of PLSV3936 and active site mutants against ulvan polysaccharide. Absorbance at 232nm was observed for reactions containing 1 µg of enzyme in 20 mM HEPES pH 7.5, 150 mM NaCl and ulvan substrate at 0.5 mg/ml concentration.

mg/ml, respectively. We attempted to crystallize these proteins in the same condition under which the native protein had been crystallized (i.e. 20% PEG 3500 as precipitant and in different salts 0.2 M ammonium acetate, 0.2 M Sodium formate, 0.2 M potassium iodide, 0.2M ammonium formate).The H264N (the presence of Zn^{2+} ion coordinated by the H264 residue were not identified at this time) and Y188F mutants precipitated, whereas H123N crystallized in PEG 3500 (15%-23%) , 0.2 M ammonium acetate and pH values between 5 and 6.5.

6.2.8 Soaking experiments of H123N mutant crystals with tetrasaccharide ulvan substrates

Mutant crystals of H123N grown in 15% PEG 3500, 0.2 M ammonium acetate, 0.1M Bis-Tris pH 5 were soaked for ~ 5 hours in a solution containing two different tetrasaccharide substrates (details as described in the LOR107 section). After soaking, the crystals were briefly washed in the cryoprotectant solution (30% ethylene glycol in mother liquor) and flash frozen in liquid nitrogen. Diffraction data was collected at the 08B1 beamline at the CLS (Grochulski et al, 2011) and processed with XDS (Kabsch, 2010). These crystals diffracted to 1.6 Å resolution and belonged to space group *C*2 with unit cell parameters a=173.6, b=71.8, c=109.3 Å and β=122.5º. There are two molecules in the asymmetric unit. The strong positive difference electron density peak in the vicinity of the active site was observed in a crystal that had been soaked in DP4-1

substrate. No such density was found in crystals soaked with DP4-2 substrate. Data collection and refinement statistics are shown in Table 6.2.1. The coordinates and structure factors have been deposited in the PDB with accession number **5UAS** for the enzyme-tetrasaccharide complex.

6.2.9 Structure of the PLSV3936 (H123N) complexed with ulvan tetrasaccharide:

The H123N mutant crystal soaked in DP4-1 tetrasaccharide diffracts to 1.6 Å resolution and the difference electron density map showed density for all four sugar residues. The uronic acid epimer in the complex structure is identified as GlcA. This suggests preferred binding of the tetrasaccharide containing GlcA over the one containing IdoA. The sulfate groups of the Rha3S residues serve as markers for placing the tetrasaccharide in the positive density and determining the directionality of the oligosaccharide. The naming of the tetrasaccharide follows the same convention described for LOR107. The reducing end of the tetrasaccharide (i.e. the Rha3S (+2 site)) sits at the entrance of the crevice, whereas the GlcA acid and a second Rha3S (+1 and -1 site) are positioned deep in the crevice. The last residue at the non-reducing end of the tetrasaccharide (-2 site), ΔUA, extends toward the surface of the crevice, making a sharp turn relative to the direction of the first three sugars **(Figure 6.2.5)**.

Three residues that were shown by mutagenesis to be essential for activity, His123, Tyr188 and Arg204, make contacts with the GlcA at the +1 site within the tetrasaccharide. The most contacts with the protein are made by the two middle sugars, GlcA (+1 site) and Rha3S (-1 site), and these two sites are determinants of substrate recognition **(Figure 6.2.6A)**. The GlcA (+1 site) contacts the protein on both sides of the sugar ring. The acidic group of GlcA forms a salt bridge with Arg204 and a hydrogen bond with $NE2^{His143}$ on one side and OH^{Tyr246} on the other side **(Figure 6.2.6B)**. This is exactly the location of the phosphate ion observed in the native structure. The arginine fulfills the role of a charge neutralizer, with the aid of His143 and Tyr246. The retaining activity of the H143N mutant indicates that the contribution of His143 on its own is less crucial. When His123 was modeled onto the asparagine sidechain in the complex structure, $NE2^{His123}$ is ~2.3 Å away from the oxygen bridging GlcA (reducing end) and Rha3S (non-reducing end) and these atoms also lie in the same plane. The modeled His123 is too close to the C5 hydrogen (NE2–HC5 hydrogen distance of 1.6 Å), indicating that the substrate is slightly shifted when bound to the active enzyme. Moreover, the H-C5 bond is nearly perpendicular to the histidine ring, an inappropriate geometry for proton abstraction. However, Tyr188 is well positioned to

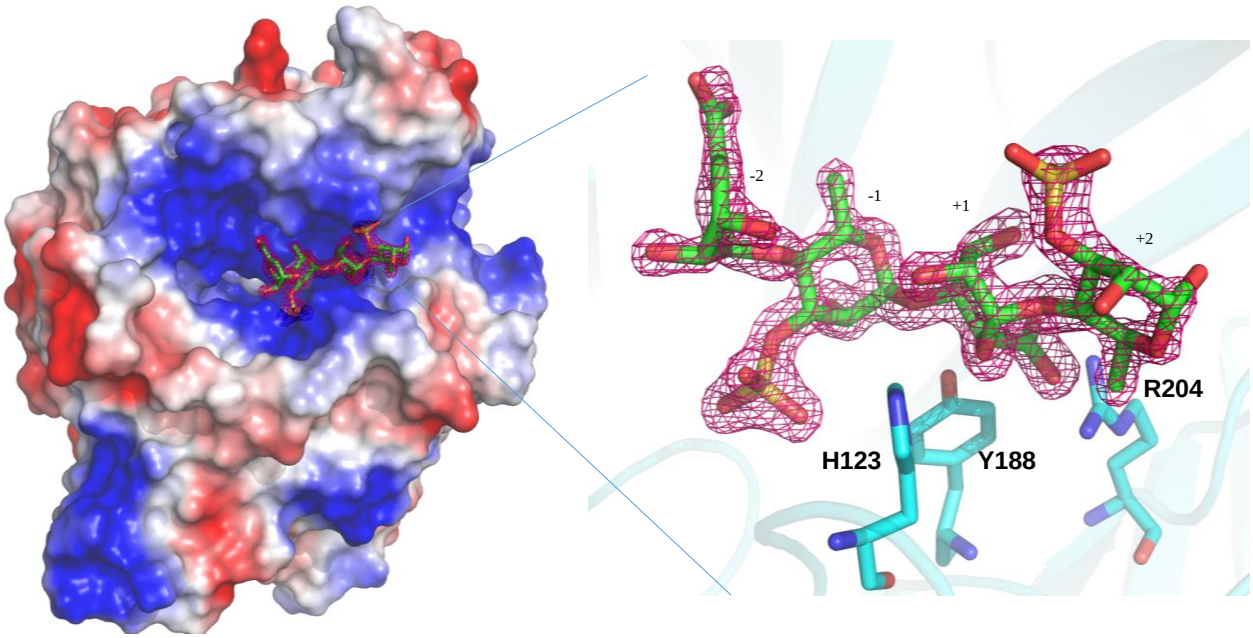

Figure 6.2:5 Location of the tetrasaccharide in the crevice.
Surface representation of PLSV3936 showing substrate in the active site crevice. The zoomed view showing the substrate and catalytic residues. Difference electron density map contoured at 3σ level around the substrate.

abstract the C5 hydrogen, with its hydroxyl group extending toward this hydrogen and with a 2.4 Å distance between OH^{Tyr188} and H-C5. There is one more hydrogen bond formed, between $O3^{GlcA}$ at the +1 site and $ND2^{His208}$.

The sulfate group of -1 site Rha3S is specifically recognized through hydrogen bonds to Asn60, Asn123, Lys125 and to Gln66 through a bridging water molecule. The O2 hydroxyl of Rha3S forms an additional hydrogen bond to Arg282 through a bridging water molecule **(Figure 6.2.6A)**. The C6 methyl group of Rha3S makes van der Waals contacts with methyl groups of Thr280, Thr355 and with the Phe329 ring. This might contribute to the binding selectivity for this sugar. The +2 site Rha3S predominantly makes van der Waals contacts with the protein, namely with the faces of His143, His158, His208 and with Ala207, while the sulfate makes hydrogen bonds with His278 and Lys354 (weak). Finally, $O2^{\Delta UA}$ at the -2 site hydrogen bonds to Asn60 and extends toward the surface.

6.2.10 Proposed catalytic mechanism

The structure of the enzyme-substrate complex, together with the enzymatic data from the mutants, allowed us to deduce the catalytic cycle and firmly identify the residues participating in catalysis. The elimination mechanism, as observed in all polysaccharide lyases, requires neutralization of the C5 acidic group on the uronic acid residue in the +1 position. In other polysaccharide lyases, this is accomplished either by a bound metal ion or an asparagine, aspartate or histidine sidechains. The neutralization of the C5 acidic group of the uronic acid in the +1 subsite is a prerequisite for making the C5 proton susceptible to attack by a base. The product analysis indicates that PLSV3936 can break the bond next to either GlcA or IdoA. Our structural data provide direct information about the binding of a substrate containing GlcA at the +1 position and delineate residues involved in its degradation. Therefore, we will discuss further the specific breakdown of this substrate.

The acidic group of the substrate is neutralized by Arg204 with some help from His143 and Tyr246. The sidechains in close proximity to the C5 proton are Tyr188 and His123. Our model suggests that a slight repositioning of the substrate in the active site of the wildtype enzyme, would give a distance of ~2-2.5 Å between the C5 proton and OH^{Tyr188} or $NE2^{His123}$. This distance is compatible with either of these residues acting as a base **(Figure 6.2.6B)**. However, the H-C5 bond would be nearly perpendicular to the histidine ring, making proton transfer to His123 less likely. Tyr188, on the other hand, is pointing directly toward H-C5 and is well poised for proton abstraction. The His123 and Tyr188 are in close proximity, with the $NE2^{His123}$ and OH^{Tyr188} of these residues are within 3.3 Å in the wildtype structure. When H123 is modelled into the H123N mutant, this distance is even shorter, allowing for a proton transfer from Tyr188 to His123. This would result in $His143^{+}$ and $Tyr188\text{-}O^{-}$, preparing Tyr188 for C5 proton abstraction. This must be followed by donation of a proton onto the bridging oxygen. His123 appears to be best positioned for this action, with the ring being ~2.5 Å away from the bridging oxygen, already in the plane of the histidine ring. Although OH^{Tyr188} is only ~3.4 Å away from this oxygen, the hydroxyl group is nearly perpendicular to the (Rha3S) C1-O-C4 (GlcA) plane and in less favorable geometry for proton transfer. Importantly, since the position of the substrate in the wildtype enzyme might be slightly different due to the fact that H123N replacement creates a small additional space for the

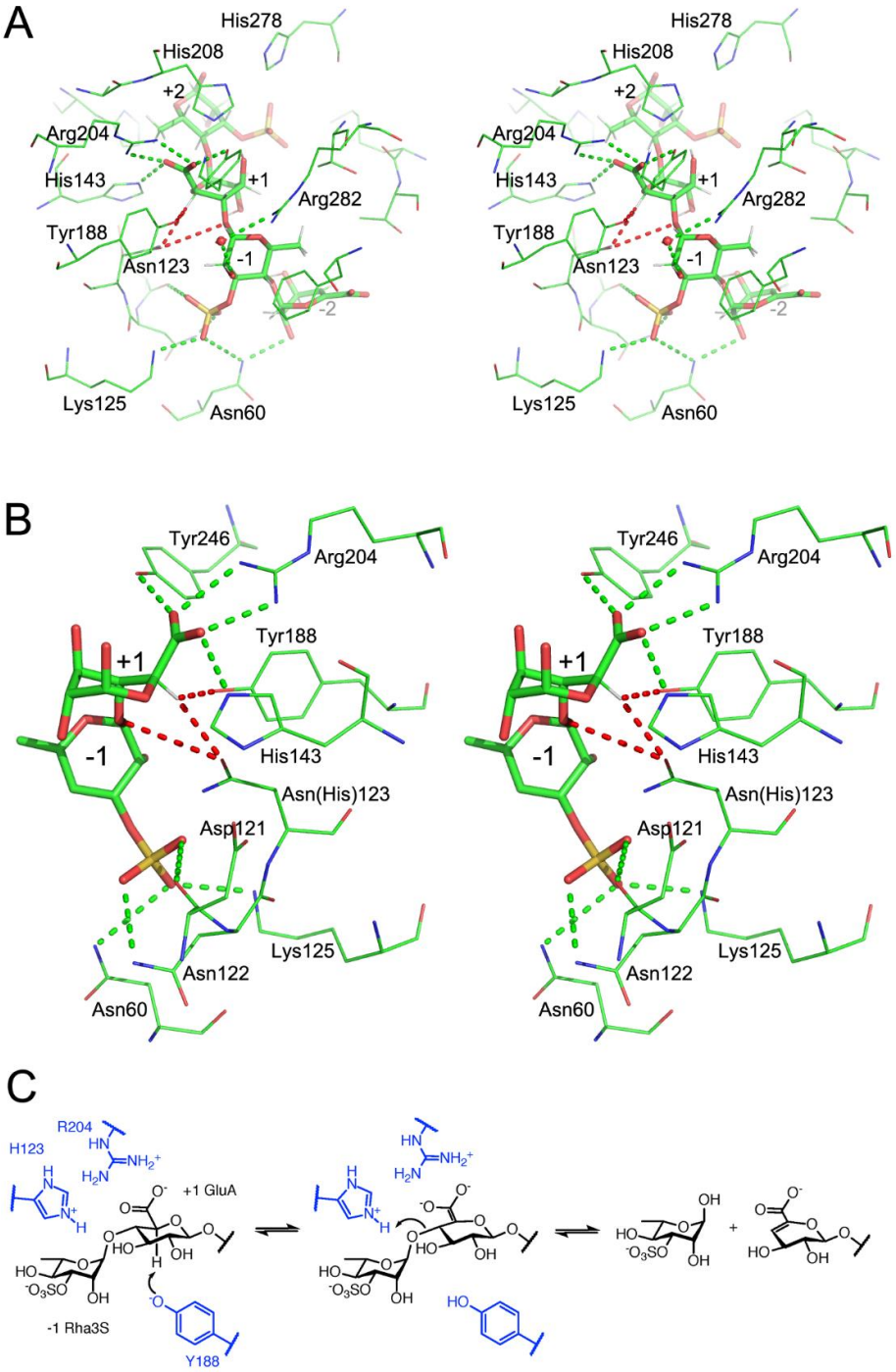

Figure 6.2:6 proposed catalytic mechanism of PLSV3936.

(A) Stereo view of the residues in close proximity to the tetrasaccharide. Green dashed lines indicate hydrogen bonds, red dashed lines indicate very close contacts; **(B)** close-up of the central sugars GlcA and Rha3S and their contacts with neighboring sidechains; **(C)** The proposed reaction mechanism of PLSV3936 with indicated roles of active site residues.

substrate to move into, we cannot be absolutely certain about the precise roles of His123 and Tyr188 from the structure alone. However, since the Y188F mutation inactivates the enzyme, we postulate that this sidechain plays a major role in catalysis and is the most likely candidate for a base. Meanwhile, His123 would then transfer one of its two protons to the bridging oxygen, with a concomitant breaking of the C4-O bond and the formation of a C4-C5 double bond. It is unclear if Tyr188 and His123 act in concert as a proton relay, or if there is a deprotonation at C5 followed by elimination. These could be either E2 or E1CB mechanisms, but we favor the latter since it is a *syn* elimination. The proposed mechanism is summarized in **(Figure 6.2.6C)**.

The H264N mutation also inactivated the enzyme, although this sidechain is not in contact with the substrate. The structure suggests a role for this residue, His264 together with His208, His278 and Cys266 are ligands of the Zn^{2+} ion. Replacement of His264 by an asparagine likely results in a loss of Zn^{2+} binding. This metal plays a structural role and its removal would destabilize the region near the active and substrate binding sites, which would inactivate the enzyme. This mutant precipitates readily during purification under the conditions where the wildtype and other mutants are soluble, supporting the notion that some structural changes have taken place.

To summarize, we solved the crystal structure of PLSV3936, a member of PL25 family in the apo state and in complex with the tetrasaccharide. This enzyme shares 14% sequence identity with PL24 family Lor107. PLSV3936 adopts the seven bladed β-propeller fold with an extended crevice on one side of the propeller. Identification of the sugar binding region, site directed mutagenesis coupled with enzyme activity measurements and the complex structure of PLSV with the dp4-1 tetrasaccharide have allowed us to delineate a possible catalytic mechanism. For this enzyme, the negative charge of the acidic group on the GlcA is neutralized by Arg204 with the help of His143 and Tyr246. Tyr188 is most likely the catalytic base while His123 likely plays the role of general acid.

Table 6.2-1 Data collection and statistics table of PLSV3936

Crystal form	Native PLSV_3936	SeMet PLSV_3936	Complex PLSV_3936
Space group	*C*2	*C*2	*C2*
a, b, c (Å) ß (°)	174.3, 72.1, 109.5 122.5	173.6, 71.8, 109.1 122.4	173.6, 71.8, 109.3 122.4
Wavelength (Å)	0.9791	0.9788	0.9871
Resolution (Å)	49.01-1.45	47.8-2.0	46.3-1.6
Observed hkl	763554	574743	1094771
# Unique hkl	194580	145561*	143664
Completeness (%)	96.2 (83.2)	98 (95.2)	96.0 (90.0)
Redundancy	3.7	3.9	7.6
Rsym	0.044 (0.71)	11.1 (0.76)	0.097 (1.35)
CC1/2	99.9 (82.2)	99.7 (69.0)	99.9 (65.5)
I/(σI)	19.15 (2.74)	11.23 (2.11)	13.75(1.5)
R_{work}	0.134		0.182
R_{free}	0.143		0.212
Wilson B ($Å^2$)	16.5		18.0
Protein	21.2 (7012)		24.0 (6956)
Ligand	36.3 (58)		28.2 (130)
Solvent	34.5 (1128)		36.4 (1128)
Favored (%)	96.8		96.2
Allowed (%)	3.1		3.8
Clash score	2.1		5.7
Bonds (Å)	0.007		0.009
Angles (°)	1.00		0.96
PDB code	5UAM		5UAS

6.3 Structural characterization of NLR48 (PL28) from *Nonlabens ulvanivorans*

NLR42 and NLR48 from *Nonlabens ulvanivorans* were the first discovered ulvan lyases. Both enzymes had poor solubility and the purified protein formed soluble aggregates. To improve solubility, the gene encoding NLR48 lacking the signal peptide ($NLR48^{(26-302)}$) was cloned into expression vectors with different tags. A shorter construct lacking a few residues at the N-terminal unstructured region (named as $NLR48^{(53-302)}$) was also made. Efforts were made to overcome the solubility issues with NLR42. Based on secondary structure predictions, one construct (NLR42-82D) lacking residues at N-terminus was made, but this truncated construct did not express well. This led us to focus only on NLR48 (From here on, the construct $NLR48^{(26-302)}$ which lacks the signal peptide is called as NLR48).

6.3.1 Expression and purification of wildtype NLR48

His tagged NLR48 and the shorter construct $NLR48^{(53-302)}$ had good expression but the purified enzyme tends to form soluble aggregates, as evidenced by SEC. His-MBP-NLR48 did not express, but the GST-NLR48 construct expressed as properly folded protein and resulted in a sample suitable for crystallization experiments. After subjecting to SEC, the monomeric peak fractions were used for crystallization trials. All the purification and SEC profiles of NLR48 are shown in **(Figure 6.3.1)**. After GST cleavage, the approximate yield of NLR48 from 6 liters of cell culture was 3 mg which was used for further experiments.

6.3.2 Crystallization of wildtype NLR48

NLR48 at a concentration of 16 mg/ml was screened for crystallization at 15ºC and 9ºC. Crystals were obtained at both 15ºC and 9ºC in 0.056M sodium phosphate monobasic, 1.344M potassium phosphate dibasic, pH 8.2 and only at 15 degrees in condition 0.49 M sodium phosphate monobasic monohydrate, 0.91 M potassium phosphate dibasic, pH 6.9. Crystals grown at 9ºC are larger than crystals grown at 15ºC. For data collection, crystals from both 9ºC and 15ºC were briefly transferred to mother liquor containing 30% ethylene glycol and flash cooled in liquid nitrogen. These native crystals diffracted to ~2.5Å resolution. Due to low sequence homology between the unknown structure and the best starting model, this structure could not be solved by molecular replacement. This led to pursue anomalous phasing using a SeMet derivative (SeMet-NLR48).

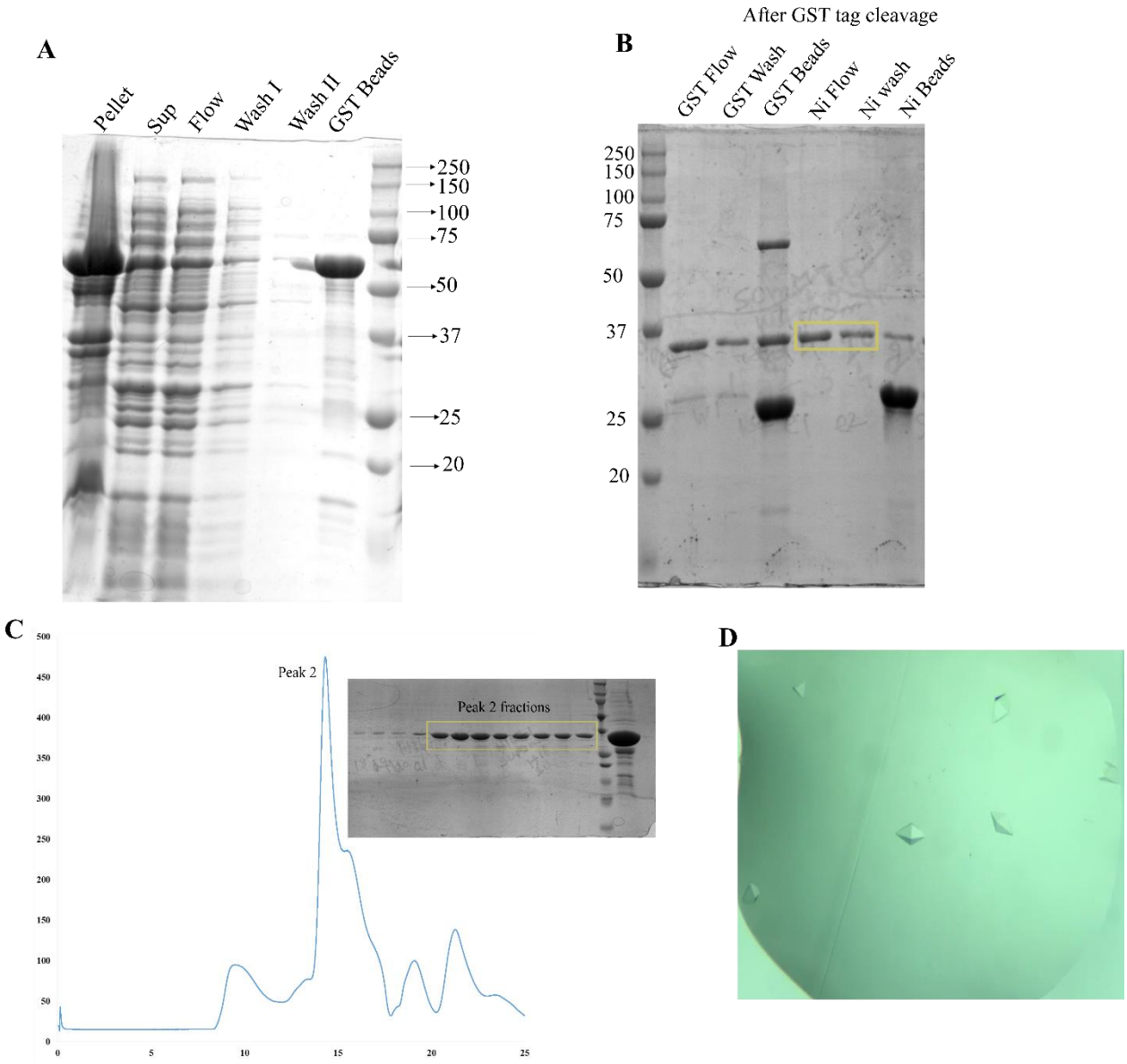

Figure 6.3:1 Purification and Crystallization of Wildtype NLR48.
(A) Affinity purification of the GST-NLR48 using Glutathione agarose resin. **(B)** Cleavage of the GST tag using His-TEV protease and the follow up purification using cobalt metal affinity resin. The NLR48 used for SEC after GST cleavage and TEV removed is highlighted in yellow box. **(C)** SEC profile of NLR48 showing the monomer behavior and the SDS-PAGE showing the purity of the peak fractions. **D)** Crystals of NLR48 obtained from 0.056 M Sodium phosphate monobasic, 1.344 M Potassium phosphate dibasic, pH 8.2 at 15° C.

6.3.3 Production, purification and crystallization of SeMet-NLR48

Selenomethionine incorporation of NLR48 was done using a methionine auxotroph mutant. SeMet-NLR48 was by purified following the same protocol used for native enzyme purification. SeMet-NLR48 was concentrated to 16 mg/ml for crystallization experiments.

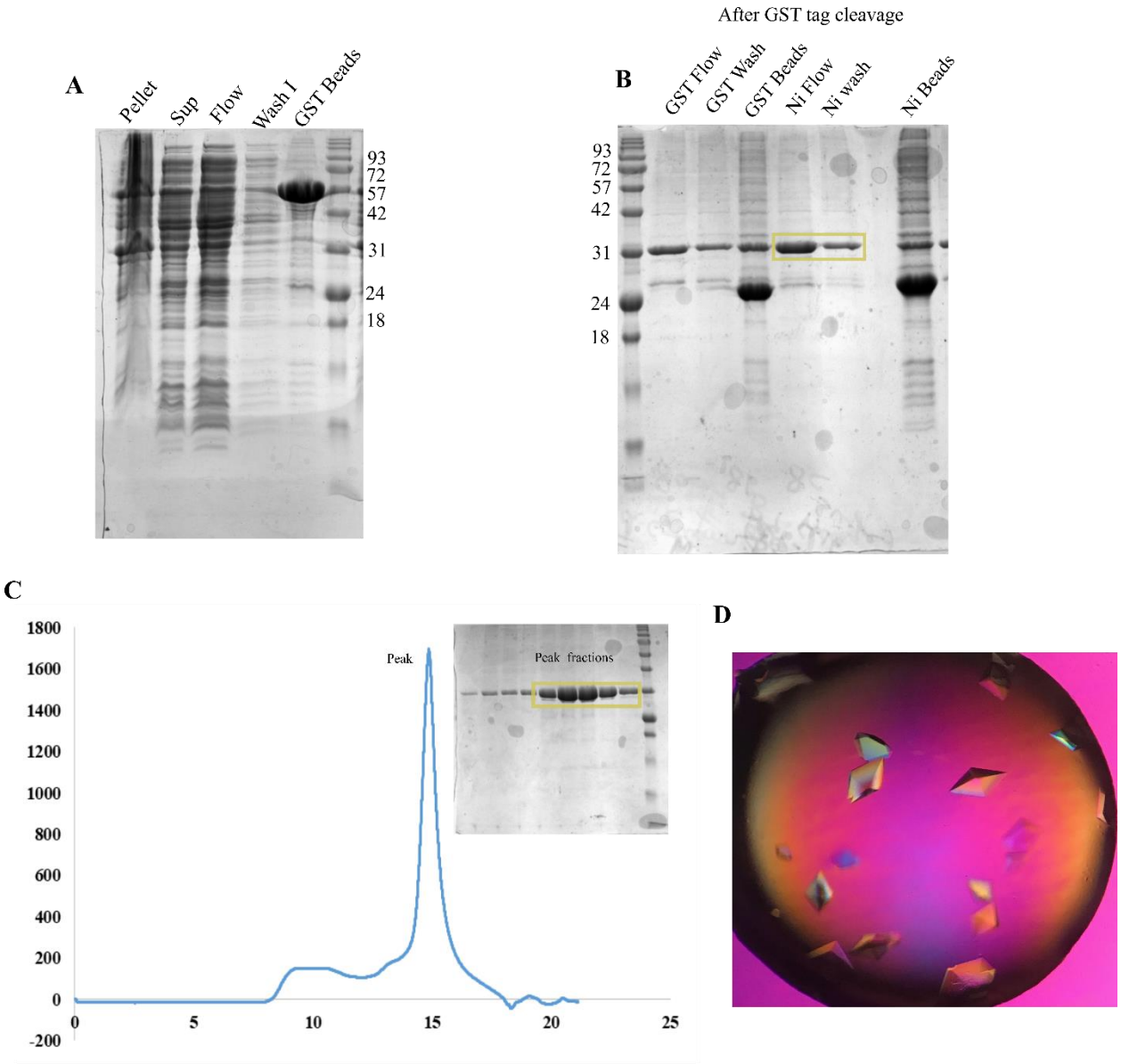

Figure 6.3:2 Purification and Crystallization of SeMet incorporated wildtype NLR48. **(A)** Affinity purification of GST-NLR48 using Glutathione agarose resin. **(B)** Cleavage of GST tag using His-TEV protease and the follow up purification using cobalt metal affinity resin. Protein after GST cleavage and TEV removal used for SEC is highlighted in yellow box. **(C)** SEC on SEC650 column showing monomeric behavior of the protein and the SDS-PAGE showing the purity of the peak fractions. **D)** Crystals of SeMet- NLR48 obtained from 0.49 M Sodium phosphate monobasic monohydrate, 0.91 M Potassium phosphate dibasic, pH 6.9 at 9^{o} C.

The crystals grown at 9^{o} C in 0.49 M sodium phosphate monobasic monohydrate, 0.91 M potassium phosphate dibasic, pH 6.9 were cryoprotected with 30% ethylene glycol and shot at the BM beamline (Fodje et al., 2014) at the CLS. Data was processed with XDS (Kabsch, 2010) and the data collection statistics are shown in Table 6.31. All the purification and SEC profiles are shown in **(Figure 6.3.2)**.

6.3.4 Structure solution and refinement of wildtype NLR48

The structure of wildtype NLR48 was solved using the autosol program (Terwilliger et al., 2009) by single anomalous dispersion method on a SeMet derivative dataset. A total of 6 Se sites were identified and the model contained residues 21-280. SeMet-NL48 crystals diffracted better than the wildtype NLR48. One of the SeMet crystals diffracted to 1.9Å. The crystal belongs to space group $P2_12_12_1$ with unit cell dimensions a=82.49, b=102.12, c=103.71. The initial model was refined against this high-resolution dataset. Refinement was done using the PHENIX refinement program (Adams et al., 2011). Manual inspection of the structure was done using COOT (Emsley et al., 2010). The final structure was validated by Molprobity (Chen et al., 2010).

6.3.5 Crystal structure of NLR48

The three-dimensional crystal structure of NLR48 was solved at 1.9Å resolution. The asymmetric unit contains two molecules related by a non-crystallographic 2-fold rotation **(Fig.6.3.3B)**. The two molecules are nearly identical and display a RMSD of 0.1Å over all Cα atoms. Each molecule covers a residue range of 21-280. Since the two molecules are identical, the following observations were made from chain A. The overall structure of NLR48 is a β-sandwich jelly roll fold with two concave beta sheets stacked together. This fold has been seen in both polysaccharide lyase (PL7, PL8, PL13, PL14, and PL20) and glycoside hydrolase (PL16) family enzymes. Each sheet has 7 antiparallel β strands (Strand A1- A7 at the inner concave and B1-B7 at the outer concave) **(Figure 6.3.3C)**. The residues Cys32 at the N terminal loop and Cys59 at the loop connecting A1 & B1 strands form a disulfide bridge. The inner β sheet (A1-A7) forms a deep cleft and electrostatic surface potential analysis (done using APBS plugin PyMol) (Baker et al., 2001) on the structure showed that this cleft is formed by positively

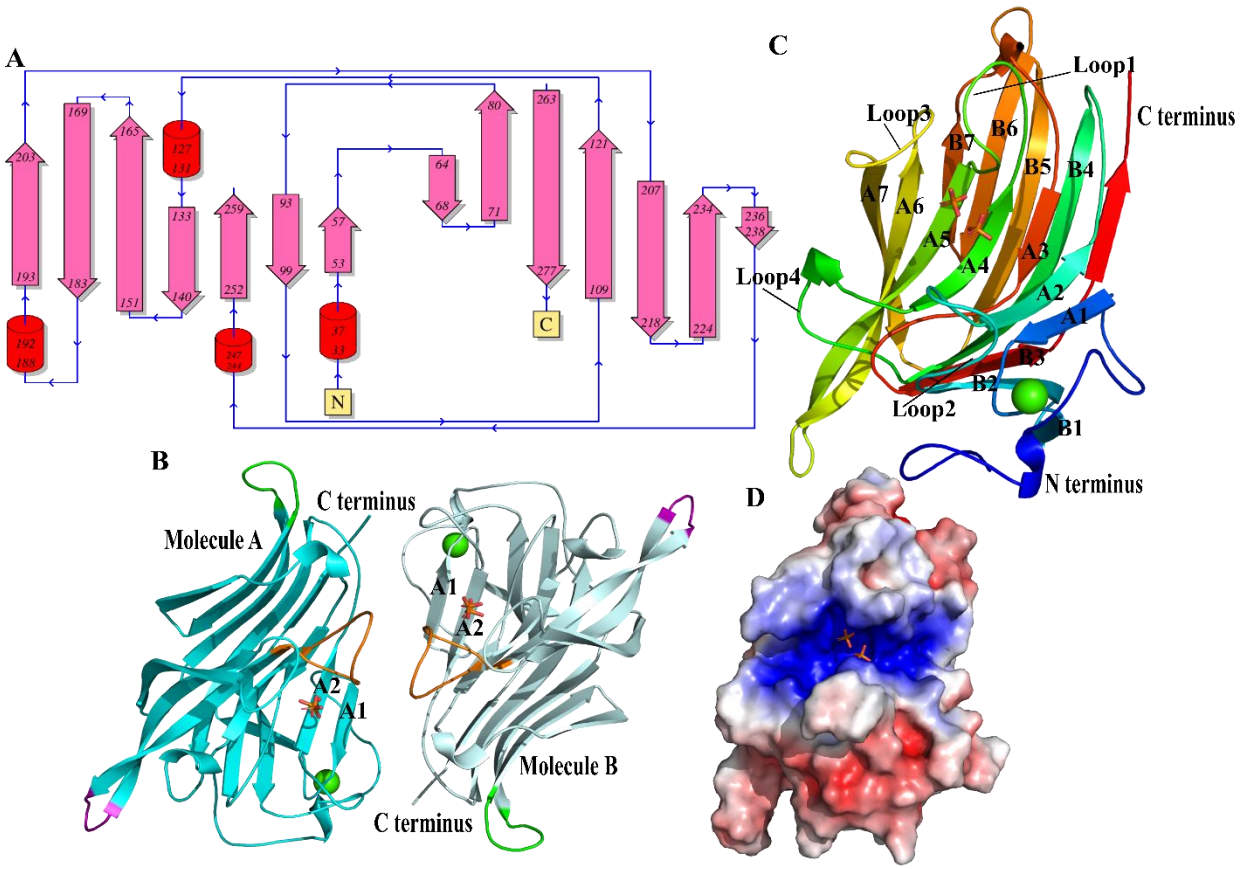

Figure 6.3:3 Structure description of NLR48.
(A) Topology diagram of NLR48 generated using PDBsum (Laskowski *et al.,* 1997). **(B)** Two molecules in the asymmetric unit related by non- crystallographic 2-fold rotation. The loops shown in green, magenta and orange corresponds to the equivalent loop in each molecule. **(C)** Schematic representation of NLR48 shown in cartoon representation and colored from N to C terminus. Calcium ion is shown as green sphere and phosphate are shown in stick representation. Strands in inner and outer concave are numbered from A1-A7 and B1-B7 respectively. Loops lining the walls of the cleft are indicated. **(D)** Electrostatic surface potential representation showing the positively charged cleft and the phosphate occupied the cleft is shown in stick representation.

charged amino acids **(Figure 6.3.3D)**. The loops connecting strands A4 & A5, B2 & A2, A6 & A7, and B4 & A4 make up the cleft wall. Both molecules have two positive difference density peaks (Fo-Fc) at the same site in the cleft, for which the scattering is higher than expected for water. Two phosphate molecules (from the crystallization solution) were modelled in the positive density and fit well in the cleft **(Figure 6.3.3D)**. This indicates that this region may be the putative substrate binding site of NLR48. Additional positive peaks were identified on both molecules at the loop region connecting strands A1 & B1. Based on the environment and the coordination geometry, these peaks are assigned as Ca^{2+} ions. This site is located away from the positively

charged cleft **(Figure 6.3.3C)** which indicates that Ca^{2+} plays a structural role in NLR48 and not in catalysis.

6.3.6 Putative substrate binding site

The Ca^{2+} ion is coordinated by pentagonal bipyramid arrangements. The equatorial ligands are backbone carbonyl oxygens of Gly^{36} (2.47Å) and Ser^{58} (2.25 Å), OD1 (2.49 Å) & OD2 (2.46 Å) of Asp^{56} and OD1 (2.46 Å) of Asn^{62}. The axial ligands are OD1 (2.31 Å) of Asn^{38} and the backbone carbonyl oxygen (2.28 Å) of Ala^{61} **(Figure 6.3.4B)**. The phosphate (PO_4^{3-}) in the positively charged cleft made many interactions with surrounding residues. PO_4^{3-} at site1 H-bonds with N2 (2.85Å) of Lys^{139}, NE2 (2.58Å) and OE1 (3.17Å) of Gln^{137}, NH1 (2.84Å) of Arg^{94}, OH (2.63Å) of Tyr^{258}. PO_4^{3-} at site2 makes H-bond with NH2 (2.72Å) of Arg^{193}, OG (2.72Å) of Ser^{148} and the backbone nitrogen (2.96 Å) of Ala^{152} **(Figure 6.3.4A)**. Based on these observations, the cleft region was considered a putative substrate binding site and all residues interacting with the PO_4^{3-} ions (Arg^{94}, Gln^{137}, Lys^{139}, Arg^{193}, and Tyr^{258}) were subjected to mutation studies.

6.3.7 Site directed mutagenesis and enzyme activity of NLR48 mutants

To obtain a structure of the enzyme substrate complex, the following mutants were made: R94N, Q137A, K139M, R193N and Y258F. All the active site mutants were expressed and purified in the same way as the native protein. The enzymatic activity of wildtype NLR48 and the various mutants on ulvan substrate was measured by monitoring the absorption at 232 nm, indicative of the formation of an unsaturated double bond in the non-reducing end. All the mutants had abolished the enzyme activity against the ulvan substrate **(Fig.6.3.4C)**.

6.3.8 Expression, purification, crystallization and structure solution of NLR48 mutants

Putative active site mutants K139M, Y258F, R94N derived from enzyme activity analysis were purified in large scale (3 liter TB for each mutant). The GST fusion constructs of the mutants were expressed and purified using the same protocol followed for wildtype NLR48. The mutant R94N was poorly expressed. This led to follow up only with K139M and Y258F mutants for crystallization trials. Purified K139M and Y258F mutants were concentrated to 17 mg/ml and set up for crystallization in 0.056 M Sodium phosphate monobasic, 1.344 M Potassium phosphate dibasic pH 8.2 and 0.49 M Sodium phosphate monobasic

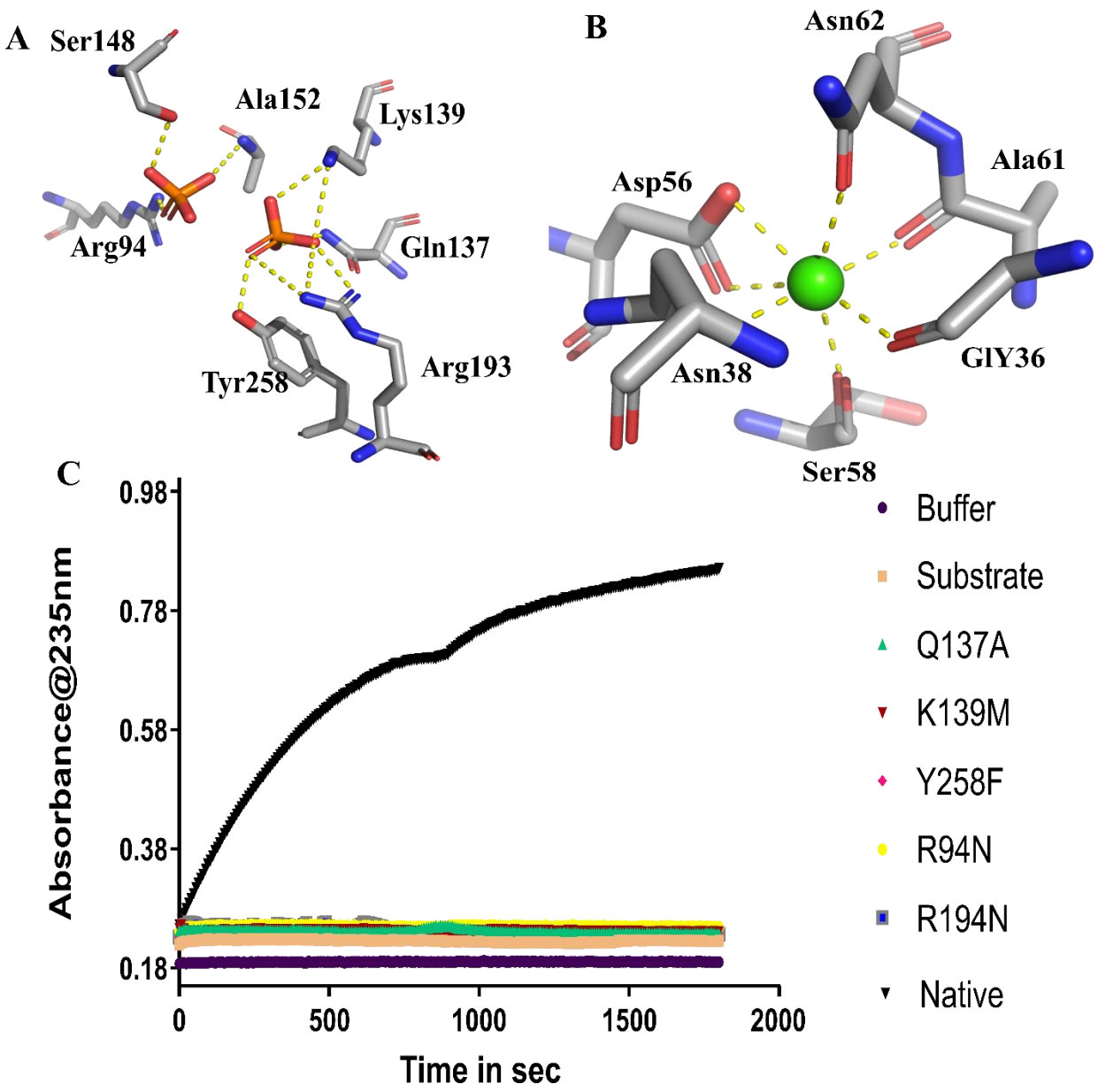

Figure 6.3:4 Putative substrate binding site.

(A) Phosphate molecules (Site I & II) in the putative active site and the surrounding residues making contact with them, all shown in stick representation. **(B)** Pentagonal bipyramidal coordination of Ca ion in wildtype NLR48. **(C)** Enzyme activity plot of wildtype and NLR48 mutants on the ulvan substrate.

monohydrate, 0.91 M Potassium phosphate dibasic, pH 6.9 at both 9^{o}C and 15^{o}C. Only the K139M mutant was crystallized and these crystals grew at both 15^{o}C and 9^{o}C in 0.49 M sodium phosphate monobasic monohydrate, 0.91 M potassium phosphate dibasic, pH 6.9. The crystals grown in the lower pH condition were soaked in tetrasaccharide substrates (DP4-1) for 2 hours and diffracted

to as low as 2.2 Å resolution. Diffraction data was collected at the ID beamline in the CLS and processed using XDS (Kabsch, 2010). The crystals belonged to space group $P2_12_12_1$ and the cell dimensions were a=82.5, b=102.7, c=103. One round of rigid body refinement was carried out on the data using the PHENIX refinement program. Positive difference density for the substrate was visible at the entrance to the active site. Manual substrate placement and further refinement were done using COOT (Emsley et al., 2010) and PHENIX, respectively (Adams et al., 2011) . Data collection and statistics are shown in Table 6.3.

6.3.9 The structure of NLR48 mutant with bound substrate

Just as in the native structure, two molecules were observed in the asymmetric unit related by a 2-fold non-crystallographic rotation. The electron density map calculated for the complex structure showed a blob of positive density in both molecules at the entrance to the cleft. The entire tetrasaccharide could be built into this density. The substrate occupies -1 to -4 subsite positions. -1, -2 and-3 subsites make many interactions with residues in the cleft and the loop region which makes up walls of cleft. Rha3S at -1 subsite interacts with Tyr^{258}, Arg^{193}, Arg^{259} and Tyr^{134}. GlcA at -2 subsite interacts with Ser^{188} and Lys^{261}. Rha3S at -3 subsite interacts with Ser^{190} **(Figure 6.3.5A & B)**. No aromatic stacking interaction is observed in the -1 to -4 subsites.

A structure comparison of the native NLR48 and the K139M mutant complex showed no significant difference in the overall fold and the substrate binding site; active site residues adopted the same conformation in both structures. The second phosphate site present in the native structure is not seen in the complex. Instead Rha3S at -1 occupies this site with a slight shift in the position. Similarly, the phosphate 1 position is also shifted slightly inside the cleft. A new positive difference density blob (not seen in wildtype NLR48 structure) is seen in the substrate binding cleft and a phosphate molecule fits well in this new site **(Figure 6.3.5C & D)**. The distance between this new site and the sugar at the -1 subsite is ~12.5Å **(Figure 6.3.5B)**. This distance can accommodate three sugar moieties, possibly +1, +2 and +3 subsites of substrate binding. Therefore, this new phosphate site might correspond to the +3 subsite vicinity of substrate. The new phosphate molecule interacts with NH2 (2.92Å) & NE (3.18Å) of Arg^{101}, NH1 (2.83Å) of Arg^{254} and OG (3.15Å) of Ser^{98} **(Figure 6.3.5B &D)**. The two molecules in the asymmetric unit appear to prevent the substrate from passing through the cleft (molecule B blocks substrate in molecule A and vice versa). Co-crystallization of the K139M mutant with the substrate did not

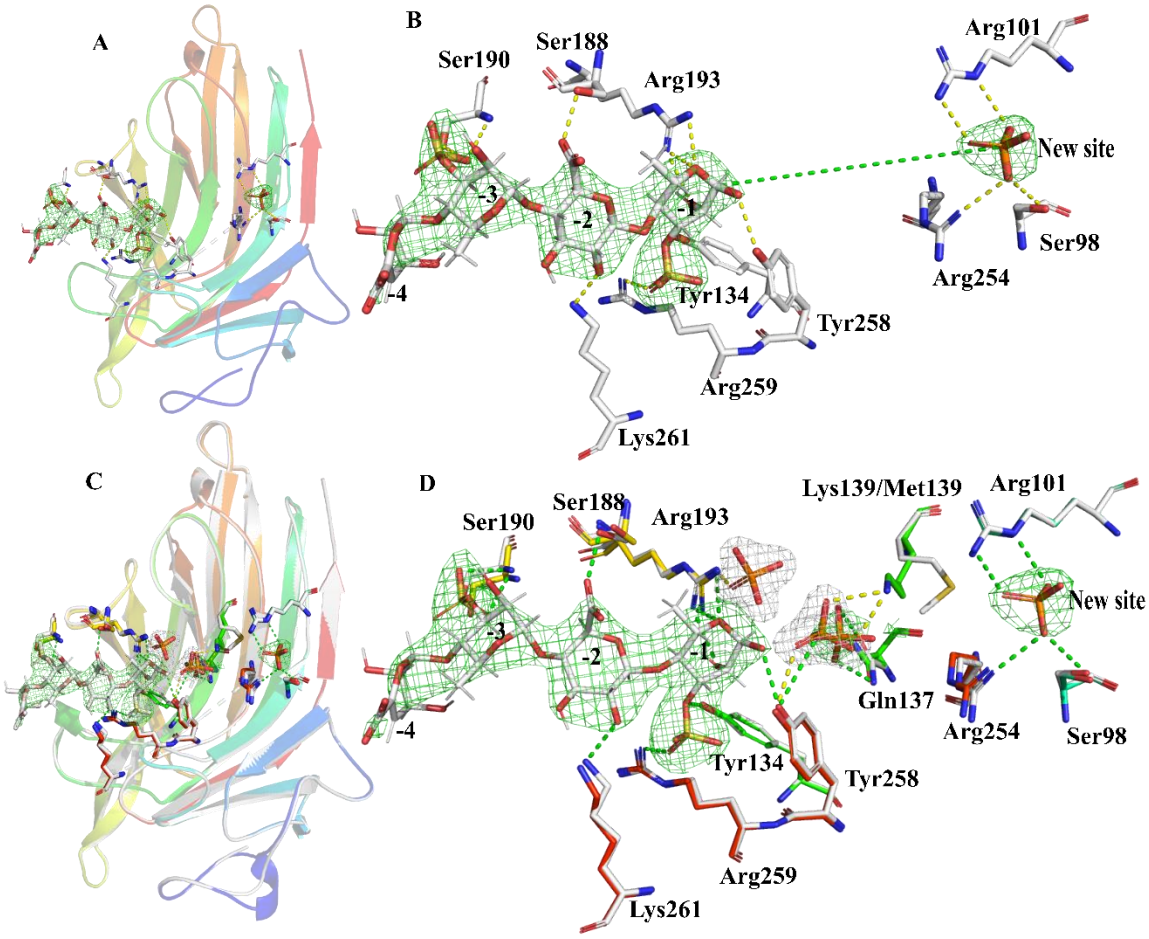

Figure 6.3:5 Complex structure of NLR48-K139M mutant.
(A) Complex structure of NLR48-K139M mutant. **(B)** The electron density (2fo-fc) map for the sugar contoured at 1.5σ. Sugars numbered from -1 to -4, the distance between the new phosphate site and the -1 substrate site is shown as green dashed lines (~12.5Å). Residues at H bonding distance to the substrate and the new phosphate site are shown in stick representation. **(C)** Structural comparison of wildtype (colored rainbow) and K139M mutants (colored white). **(D)** The electron density (2fo-fc) map for the sugar contoured at 1.5σ. -1 subsite occupies the vicinity of phosphate site 2 of wildtype with small shift. Phosphate at site I in K139M complex has a slight shift moved inner because of the substrate binding. This may correspond to +1 subsite of sugar. The space between the new site and the -1 subsite is which can fit 3 sugar moieties.

yield any crystals in this condition. This outcome could be rationalized by suggesting that the presence of substrate disrupts the crystal contacts at the dimer interface. This may also explain why the substrate docked at the active site entrance spanned the -1 to -4 subsites. Also, the high concentration of phosphate in the crystallization condition might effectively block the center of the cleft from the incoming tetrasaccharide and prevent it from occupying the +1 subsite.

6.3.10 Comparison with other β-jelly roll fold enzymes

The glove-like β-jelly roll scaffold has been reported in many polysaccharide lyase (PL7, PL14, PL18, PL13 and PL20) families (M. L. Garron and Cygler, 2014). Among these PL7, 14 and 18 have enzymes which degrade alginate. PL13 has heparin lyase I which degrades the GAG heparin. PL20 has an enzyme which cleaves Poly-GlcA substrates. Interestingly, all these enzymes share very little sequence identity (~<14%) with NLR48. The number of β-strands in the inner and outer β-sheet varies but the core structures of the enzymes superpose well with the core of NLR48. The putative catalytic site residues of NLR48 are structurally conserved among those lyases. Differences appear in the loop region **(Figure 6.3.6)**. In PL7 alginate lyase the lid loop mechanism or loop motion is important in substrate recognition and binding (Ogura et al., 2008). Also in HepI from PL13 family the mobility of lid loop and the presence of a thump domain is important for substrate binding (Han et al., 2009). In NLR48 there is no indication of loop mobility because the electron density of these loops are clearly defined. This fold is also seen in glycoside hydrolase GH16 family. GH16-Kappa carrageenases, enzymes degrading marine sulfated polysaccharide carrageenans have a β-sandwich domain with a tunnel shaped active site (Matard-Mann et al., 2017).

6.3.11 Possible catalytic mechanism

A complex structure with substrate has been reported for the PL7 family alginate lyase A1-II' (Ogura et al., 2008) and PL13 family heparin lyase (HepI) (Han et al., 2009). In both enzymes, a His/Tyr mediated reaction mechanism is observed. Therefore, to understand the role of the residues in the cleft, these structures (NLR48, HepI/substrate complex and A1-II'/substrate complex) were superimposed based on the backbone atoms of their conserved residues. This superposition represents the local environment of the active site. In this superposition the +1 uronate is the only sugar that has similar position and orientation among the structures. The +2 and -1 sugars are oriented differently. Nevertheless, the O1 of Rha3S in the NLR48 complex is in nearly the same place as the corresponding anomeric O atoms in -1 sugars of the other two complexes. It is therefore very likely that the +1 GlcA/IdoA sugar is placed in the same way as in the other two lyases. In such a case, Gln137 can neutralize the acidic group on the +1 sugar and forces the C5 deprotonation. Tyr258 is in the vicinity and its hydroxyl group is pointing toward the bridging oxygen. With GlcA in the +1 site the hydroxyl group of Tyr258 is in close proximity

6.3 Structure and catalytic mechanism of NLR48

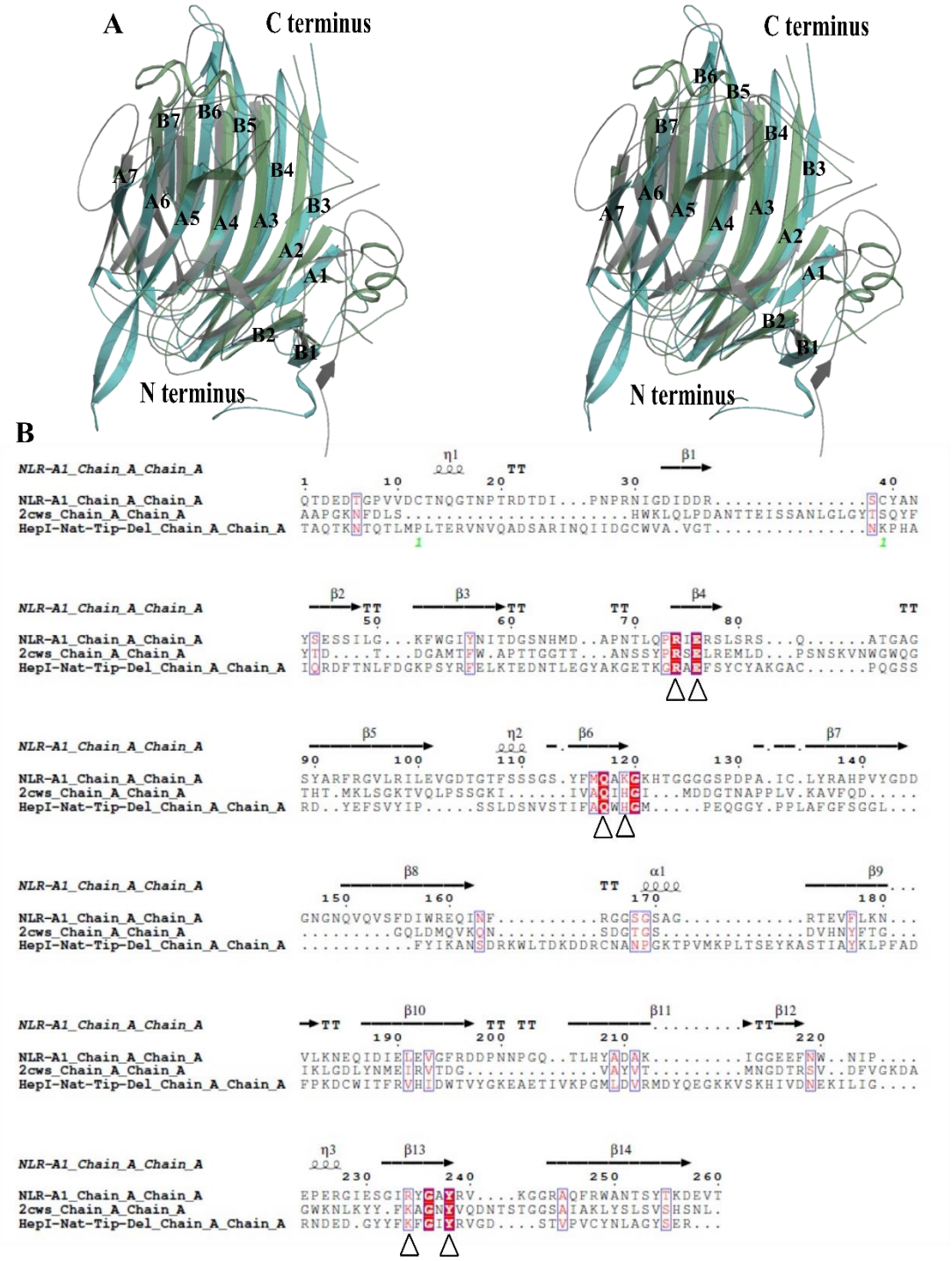

Figure 6.3:6 Structural comparison of NLR48 with other β jelly roll fold enzymes
(A) Cross eyed stereo view of NLR48 (colored cyan), PL7 family alginate lyase PDB code: 2CWS (colored green) and PL13 family HepI PDB code: 3IKW (colored grey). The inner and outer concave β strands are numbered based on NLR48 structure from A1-A7 and B1-B7, respectively. **(B)** Sequence alignment based on structure superposition, conserved residues used to align the active site (R94, E96, Q137, K139, R254 and Y258) are marked with triangle.

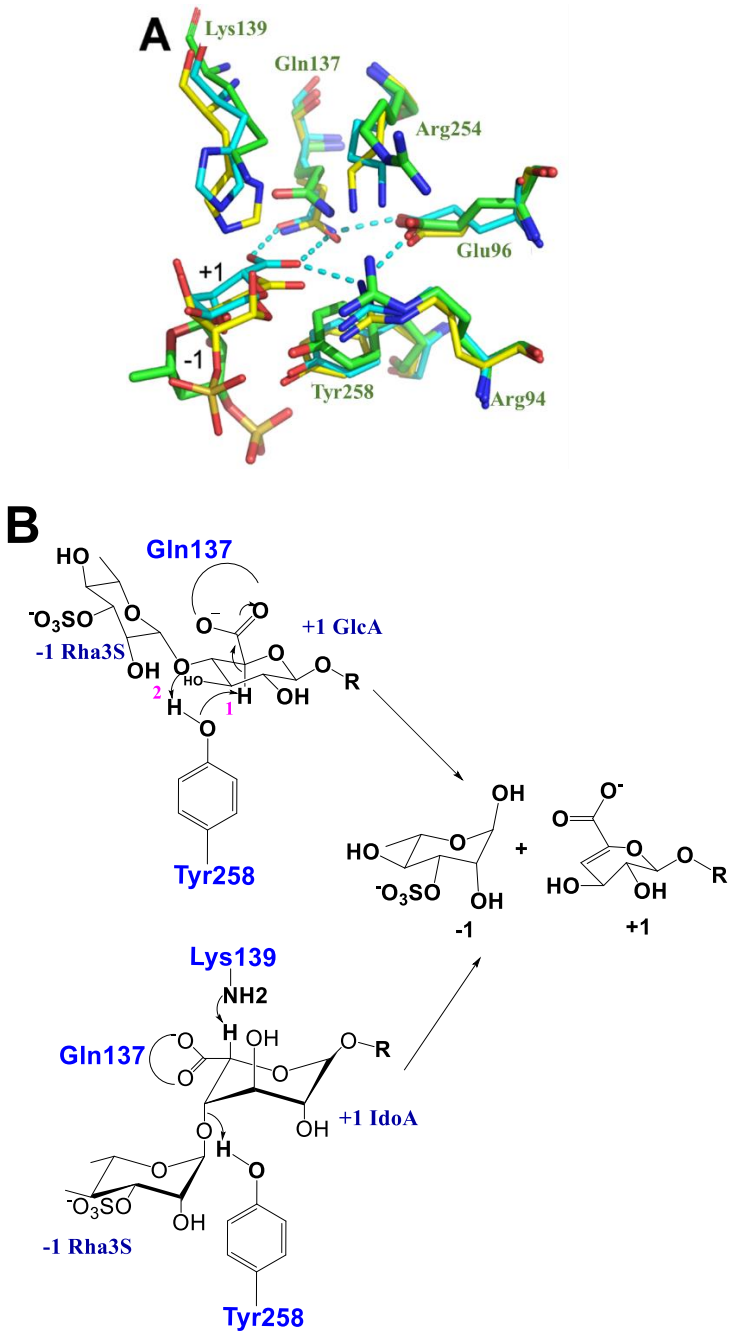

Figure 6.3:7 Putative catalytic mechanism of NLR48.

(A) Superposition of enzyme-oligosaccharide complexes based on the backbone of conserved residues in the cleft (Arg94, Glu96, Gln137, Arg254 and Tyr258 and their equivalences in other lyases. NLR48, PL7 (2ZAA) and PL13 (2INA) are colored green, cyan and yellow respectively. Only the +1 sugar in PL7 and PL13 and -1 subsite (Rha3S) of NLR48 is shown for clarity. The hydrogen bond network between the acidic group of +1 sugar, Gln, Arg and Glu is shown with dashed lines **(B)** (Top panel) Tyr258 act as a base and acid for the cleavage near GlcA, while for IdoA the base is Lys139 and the acid is Tyr258 (bottom panel).

to the C5 proton. Therefore Tyr258 could perform the role of both proton abstraction and donation. In the case of a IdoA, the proton is on the opposite side of the sugar ring to Tyr258. The only sidechain close to this proton is Lys139. There is a histidine in an equivalent position in PL13 lyases that cleave next to IdoA and in PL7 lyases that cleave next to guluronate, the same epimer as IdoA. However, in PL20 lyases, which cleave next to GlcA, the C5 proton is on the opposite side facing the tyrosine and the Lys139 position is taken by an isoleucine. Since NRL48 can likely cleave the bond next to either epimer of uronic acid like its homolog NRL42, Lys139 can act as an acid for proton abstraction. The role of a lysine residue as a proton acceptor is not unusual and seen in other PL families with different folds, namely PL3, PL9, PL6 and PL11. Overall, both the fold and the catalytic machinery are conserved between families PL28, PL7, PL13, PL18 and PL20.

To summarize, we solved the crystal structure of NLR48, a member of PL28 family in apo state and complex with the tetrasaccharide. NLR48 adopts the β-jelly roll fold with an extended substrate binding site. A complex structure of NLR48 with substrate and comparison with structural homologs help us to delineate the possible catalytic mechanism. The negative charge of the acidic group on the GlcA/IdoA is neutralized by Gln137 with help from Arg94 and Glu96. For the cleavage near GlcA Tyr281 act as a base and acid, while for IdoA the base is Lys139 and the acid is Tyr258.

Table 6.3-1 Data collection and statistics table of NLR48

Crystal form	NLR48(SeMet)	NLR48(K139M)-tetrasaccharide
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2_1
a, b, c (Å), ß (º)	82.5, 102.1, 103.7	82.5, 102.7, 103.0
Wavelength (Å)	0.98013	0.97949
Resolution (Å)	50-1.9	50-2.2
Observed hkl	1,001,093	432,911
# Unique hkl	131,854	44,398
Completeness (%)	99.8 (97.7)	99.8 (97)
Redundancy	7.6	9.8
Rsym	0.143 (1.25)	0.10 (1.16)
I/(σI)	10.01 (1.82)	14.86 (2.0)
Refinement statistics		
R_{work}	0.169	0.189
R_{free}	0.201	0.220
Wilson B ($Å^2$)	26.28	49.57
Ramachandran plot (%)		
Favored	96.57	95.35
Allowed	2.67	3.88
Outliers	0.76	0.78
R.m.s. deviation		
Bonds (Å)	0.015	0.008
Angles (°)	1.214	1.091
PDB code	6D2C	6D3U

Chapter 7

---

7. Discussion

The main goal of the research described in this thesis is to expand our structural and mechanistic understanding of the enzymes catalysing the breakdown of naturally occurring polysaccharides *via* the β-elimination mechanism. Polysaccharides that can be degraded by this mechanism contain a sugar with a carboxylic group at the C5 position. General steps involved in the β-elimination mechanism are: (1) neutralization of the acidic group, (2) abstraction of the C5 proton and (3) transfer of a proton to the leaving group (Gacesa, 1987). The main classes of polysaccharides that are substrates for polysaccharide lyases (PLs) include pectin/pectate, rhamnogalacturonan, heparin, heparan sulfate, heparosan, chondroitin sulfate, dermatan sulfate, alginate xanthan, glucuronan and ulvan. The PLs are classified based on their sequence similarities into 27 families (CAZy database). Five of these families were added to the database only recently (PL24-PL28). At the onset of this work there were no structural data for the PL12 family. Therefore, HepIII of the PL12 family from *Bacteroides thetaiotaomicron* became the subject of my research. Subsequently, three new PL families with specificity for ulvan were identified (PL24, PL25 and PL28) and the representative lyases from each of these families became additional subjects of my studies. Polysaccharides are commercially important and are utilized in food, pharmaceutical, cosmetic, biomaterial and other industries. Polysaccharide lyases are useful tools to process polysaccharides and tailor them to specific applications.

Four polysaccharide lyases have been structurally and mechanistically characterized in this research, namely HepIII, LOR107, PLSV3936 and NLR48. These enzymes provide the structural templates for lyases in four PL families.

Structural comparison of HepIII with other GAG lyases

HepIII has an $(\alpha/\alpha)_5$ toroid + β-sandwich fold. This fold had been observed in other GAG lyase families, namely PL8 (Hyl, ChonAC, ChonABC) and PL21 (HepII). Crystallographic investigation to understand the molecular basis of HepIII-substrate recognition was unsuccessful. However, the active site residues (Tyr301, Asn247, and His431) were deduced from structural comparison with the closest structural homolog, HepII. The participation of these residues in catalysis/substrate binding was further confirmed by mutation and the lack of enzymatic activity of the mutants on its natural substrate, heparan sulfate (HS).

HepII can degrade HP and HS by cleaving next to both GlcA and IdoA epimers. The catalytic residues involved in GlcA degradation are Glu205 (neutralizer) and Tyr257 (base/acid) whereas for IdoA degradation, the *anti* elimination, involves Glu205 (neutralizer), His202 (base) and Tyr257 (acid). Based on structural comparison with HepII and the known specificity of HepIII for the GlcA epimer, it is most likely that Tyr301 performs the roles of both catalytic base and acid during *syn* elimination. Neutralization of the carboxylic group in the +1 site to force the C5 deprotonation is most likely performed by Asn247. The role of Asn247 as a neutralizer is further supported by the structural comparison of HepIII with chon AC and Hyl. In these proteins, charge neutralization is accomplished by the asparagine sidechain (Li et al, 2000; Lunin et al, 2004).

The structures of the three lyases from PL12 are currently known and none of them were crystallized with the substrate. A comparison of their structures showed that the individual domains of the proteins overlap well, whereas their relative orientations differ in each enzyme. Our molecular dynamics simulation *via* NMA calculated large-scale movements with the hinge point at the bottom of second helix in the third α-helical hairpin; near Asn281 at the N-terminal domain. The movement trajectories of these proteins intersected in one conformation that represents a semi-closed state. We postulated that this is the most likely active conformation of these lyases and the conformations observed in the crystal are more stable but not catalytically compatible. The shift to an active conformation is likely driven by the energy derived from substrate binding and may require a longer polysaccharide. It might be that this conformational change is stimulated more easily by a longer HS rather than by a tetra- or hexasaccharide that was used for co-crystallization attempts.

Heparinase enzymes in Gut microbiome

The microbes residing in the gut depend on the dietary and host-derived glycans for their nutrient source. *Bacteroides thetaiotaomicron,* one of the major members of the gut-residing microbes, depend on HS and HP as their major nutrient source. The heterogeneity of HS and HP in terms of the degree of polymerization (DP), sulfation pattern and the uronate epimerization increases the complexity of these polymers and makes their degradation more challenging. Hence, the *B. thetaiotaomicron* genome contains multiple genes related to HS digestion. They are organized as HS utilization loci that encode a repertoire of enzymes to degrade the host HP and HS. Recent research on the HS PULs in *Bacteroides thetaiotaomicron* proposed a model for HS degradation. The key findings are: 1) HepIII, the product of *bt*4662 (PDB code: 4FNV), is identified as a surface lyase, with a highest degree of degradation on HS and desulfated HP. Bt4662 was also identified as an endo-lyase and is active on oligosaccharides containing a minimum of 10 monomer units. This is in accordance with our finding that the catalytically competent active state of HepIII is driven by the binding of longer substrates. 2) In addition, HepIII (product of *bt*4657, PDB code: 5JMD) was identified as an endo-acting periplasmic lyase with some degree of processivity. Consistent with our substrate specificity assay, 5JMD shows highest activity against HS (less sulfated). 3) PL13 family HepI, the gene product of *bt*4675, showed higher activity against HP and was inactive against desulfated Hep, which suggest that the sulfate group is required for its activity. Overall our findings on the substrate recognition and specificity of heparinase enzymes through crystallography help in understanding the big picture of the HS degradation system (Cartmell et al., 2017).

Ulvan lyase

Ulvan degrading lyases have been identified and characterized only recently. This discovery led to the addition of new families in the CAZy database namely, PL24, PL25 and PL28. We have determined the first crystal structure of a member of each family; LOR107 (PL24), PLSV3936 (PL25) and NLR48 (PL28). LOR107 and PLSV3936 adopt a seven bladed β- propeller fold, whereas NLR48 has a β-jelly roll fold. All three enzymes utilize the His/Tyr catalytic mechanism with slight modifications.

Comparison of LOR107 and PLSV3936

LOR107 and PLSV3936 adopt the same fold, despite sharing only ~14% sequence identity. The two structures can be superimposed with a RMSD of 1.73 Å for 188 Cα atoms (out of ~440 common). The parts of the structures that are similar include β-strands of the propeller, while the loops differ in their conformations **(Figure 7.1A)**. Despite adopting the same fold, the two structures display several features that differentiate them: **(1)** in the structural superposition, the β-strands of the blades align well but the loops diverge; they have different lengths and conformations, particularly the long loops at the top of the propeller. **(Figure 7.1A)**. **(2)** While both enzymes contain structural metal ions, LOR107 has two Ca^{2+} ions that are distant from the active site while PLSV3936 contains a Zn^{2+} ion that is located closer to the active site and likely rigidifies it. **3)** The residues crucial for LOR107 activity are divided into two clusters, whereas the activity of PLSV3936 depends on residues forming one cluster. The LOR107 residues from cluster I, His146, His167, Thr242, Tyr243 and Arg259 have structural equivalents His123, His143, Thr187, Tyr188, and Arg204 in PLSV3936, respectively. In the structural superposition, the blades 1-3 and 7 of the propeller in both proteins align well while blades 4-5 in LOR107 are shifted more outside, resulting in widening of the substrate binding canyon **(Figure 7.1 B)**. **(4)** As a result, cluster II residues in LOR107 (Asn263, Tyr303, Arg320 and Tyr330) are shifted slightly outside. Moreover, in cluster II of LOR107, only Tyr303 has a conserved counterpart (Tyr246 in PLSV3936). The residue that corresponds spatially to Arg320 in LOR107 is His264 in PLSV3936 **(Figure 7.1 C)**. His264 in PLSV3936 mediates the Zn^{2+} ion and mutating this residue leads to aggregation of the protein. Whereas Arg320 in LOR107 is not involved in metal coordination but plays a key role in substrate binding. **(5)** Another notable difference is the essential role of His167 (cluster I) in the catalysis of LOR107, while mutation of the structurally equivalent residue His143 in PLSV3936 does not affect the enzyme activity.

Substrate binding in LOR107 and PLSV3936

The substrate binding in LOR107 is captured as a two stage binding event in the crystal structures. A comparison of the substrate binding in PLSV3936 with LOR107 (R320N mutant complex structure) shows that the tetrasaccharide binds in a similar region relative to the catalytic residues. The -2 and -1 sugars take on the same conformations. However, the +1 and +2 sugars show different orientations. The +1 GlcA are in slightly different orientations relative to the active

site residues in these two lyases but the orientation of the +2 Rha3S differ significantly; the rings are flipped by ~180^{o} and the sulfate groups point in opposite directions in these two binding sites **(Figure 7.1 D)**, indicating that the '+' sugars follow different paths along the enzyme surfaces. The small difference in the orientation of the +1 GlcA indicates some plasticity in the His/Tyr active site that can accommodate changes in substrate orientation and maintain efficient bond cleavage.

Biochemical studies showed that LOR107 cleaves specifically after GlcA (+1 site) and PLSV3936 can cleave after GlcA/IdoA (+1 site) within the ulvan polysaccharide. The crystal structures of PLSV3936 was captured with GlcA at the +1 site. However, the mutations of Arg204 (neutralizer), Tyr188 (base) and His123 (acid) in PLSV3936 completely inactivate the enzyme. This suggests that these sidechains could be involved in degrading both uronic acid epimers. However, in IdoA, the C5 proton is on the opposite side of the sugar ring to Tyr188. There are two residues, Tyr246 and His208, facing the C5 proton of IdoA **(Figure 7.1 F)**. In that, His208 is the Zn^{2+} coordination ligand. At this point, the mechanism of cleavage next to IdoA by PLSV3936 is less clear. A detailed understanding of the IdoA-containing ulvan degradation mechanism requires structure determination of PLSV3936 in complex with the ΔUA-Rha3S-IduA-Rha3S substrate. However, positioning IdoA at the +1 site in LOR107 shows that Tyr303 would point toward the C5 proton **(Figure 7.1 E)** and mutation of Tyr303 to Phe does not affect the enzyme activity. This structural data, along with mutation studies and enzyme activity assays, support the substrate specificity of LOR107 for GlcA epimer.

Comparison with NLR48

Despite its different fold, NLR48 shares similarities with the PL24 and PL25 lyases in their catalytic mechanisms. The biochemical characterization of NLR48 finds that NLR48 can cleave next to both uronic acid epimers (Colle et al., 2011). In accordance with that result, mutation of Lys139 (potential base for IdoA cleavage) and Tyr258 (potential acid for IdoA as well potential acid/base for GlcA) inactivates the enzyme. Therefore, the catalytic site of NLR48 might be similar to PLSV3936, provided that Tyr256 is involved in catalysis. Nevertheless, all three lyase enzymes utilize the His/Tyr mechanism. However, neutralization of the C5 acidic group is achieved by different means. NLR48 uses a glutamine that faces the carboxyl and forms two hydrogen bonds with it. A neighboring Glu96 and Arg94 residue assists in neutralization. In PL24 and 25 enzymes,

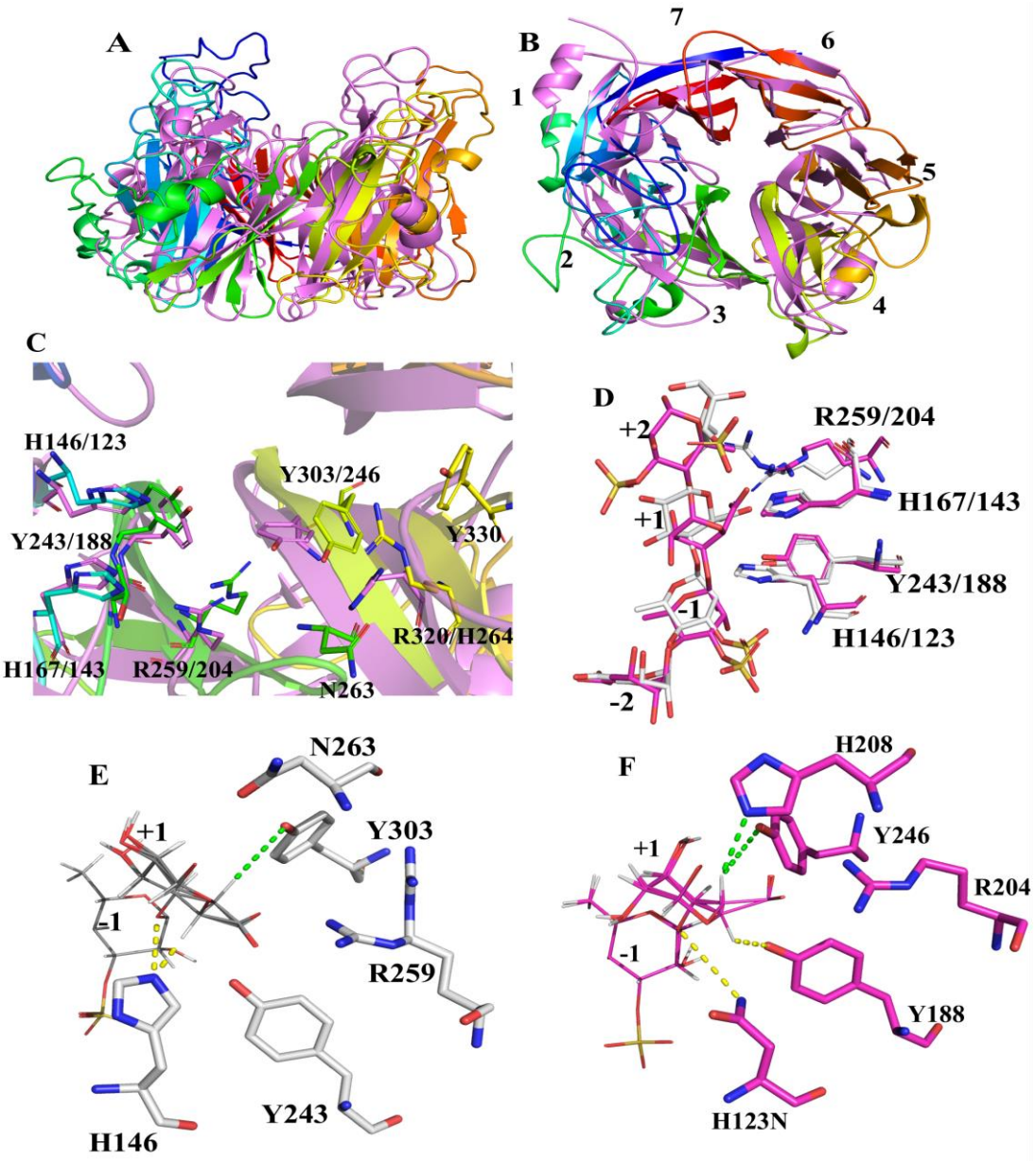

Figure 7:1 Structure comparison of LOR107 and PLSV3936.

(A) Cartoon representation of LOR107 (colored as rainbow) and PLSV3936 (colored magenta) superposition, the divergence of loops on the top of propeller. **(B)** The same superposition as A but rotated 90^{o} along Y axis, showing the widening of barrel in LOR107 by the strand deviation in blade 4 and 5 (the loops are smoothened for better view). **(C)** Zoomed view of active site vicinity showing the equivalent residues in both structures. **(D)** Substrate binding and catalytic residue in LOR107 (colored white) and PLSV3936 (colored magenta). **(E, F)** IdoA at +1 site in LOR107 and PLSV3936 respectively, the green dashed line showing the interaction of C5 proton and the

potential catalytic residues, the yellow dashed line showing the catalytic residues of GlcA cleavage.

the neutralization of the +1 sugar acidic group is accomplished by the formation of a salt bridge with an arginine residue.

Comparison with other PL family enzymes

The β-propeller fold has been observed in two other PL families: a seven-bladed β-propeller in PL22, oligogalacturonan lyases (OGL) from *Yersinia enterocolitica* (PDB code: 3PE7) (Abbott et al., 2010) and an eight-bladed β-propeller in PL11, rhamnogalacturonan lyases (RGL) (Ochiai et al., 2009). The structural comparison of PL22 OGL and PLSV3936 shows that the two enzymes have roughly the same barrel diameter and that PLSV3936 has extended loops that form the substrate binding site **(Figure 7.2B)**. The structure based alignment of PLSV3936 with 3PE7 shows only ~5% sequence identity and blade *n* of 3PE7 superimposes with blade *n*+1 of PLSV3936.

The structure of LOR107 and OGL superimpose with a RMSD of 1.7 Å for 95 Cα atoms. This superposition covers only four blades. In addition, the orientation of β-strands relative to the central axis of the propeller differs between the two structures. This leads to differences in the diameter of the propellers, with those of LOR107 being wider than OGL **(Figure 7.2A)**.

There is currently no structure of OGL in complex with a substrate. Consequently, only a putative catalytic mechanism has been proposed for this enzyme. The OGL mechanism differs from that of ulvan lyase, in that it likely uses a metal ion (Mn^{2+}) mediated β elimination reaction mechanism. His242 was proposed to act as a base and two other residues (Arg217 and His211) were also identified to play a role in catalysis. These putative residues are located on the 4th and 5th blade, whereas the catalytically important residues in LOR107 and PLSV3936 are in the 2nd and 3rd blades **(Figure 7.2C)**. Despite having the same fold, the putative active site residues in OGL are in a different structural context than LOR107 and PLSV3936. The presumed catalytic residues cannot be superimposed and the participation of Mn^{2+} in OGL catalysis suggests a distant evolutionary relationship (if one exists at all).

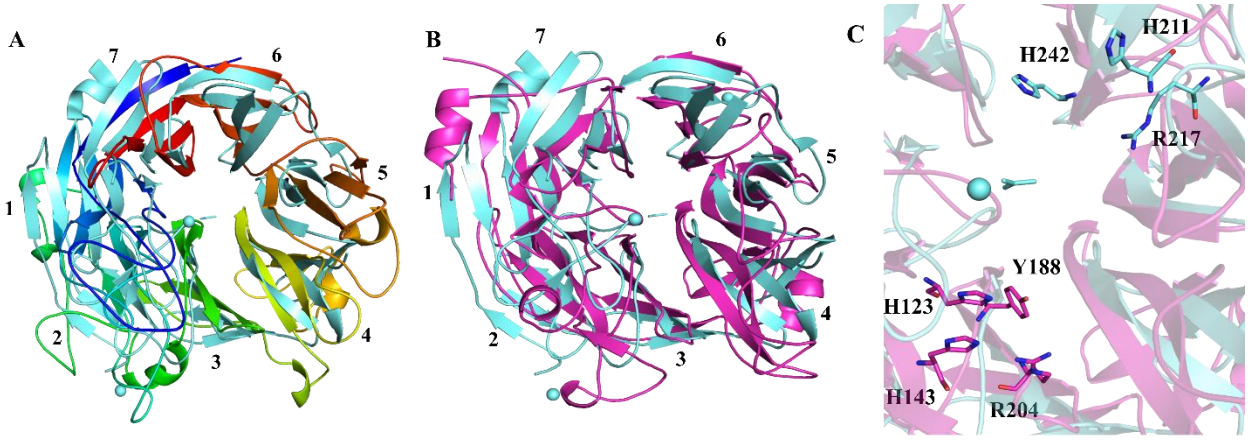

Figure 7:2 Structure comparison of LOR107, PLSV3936 and PL22 OGL
(A) Cartoon representation of LOR107 (colored as rainbow) and OGL (colored cyan) superposition showing the widening of propeller in LOR107 along blade 4 and 5 (naming of blade based on LOR107). **(B)** Cartoon representation of PLSV3936 (colored magenta) and OGL (colored cyan) (naming of blade based on PLSV3936) superposition showing the propeller. In both A & B, numbering of blade is based on ulvan lyase. **(C)** Zoomed view of active site vicinity of PLSV3936 and OGL showing the catalytic residues are in different blades.

Convergence of the catalytic site

The β-elimination mechanism employing His/Tyr catalytic apparatus is one of the two catalytic mechanisms observed among PL enzymes. This favours the *syn* and/or *anti* β-elimination reaction mechanism. All three ulvan lyases adopt the His/Tyr reaction mechanism with slight modifications. Specifically, LOR107 and PLSV3936 appears to use an Arg residue as neutralizer. Most PL family enzymes that utilize this mechanism were reported to use Asn, Asp, Glu or Gln as a neutralizer of the C5-acidic group, to force its deprotonation. Only in the PL15 family (an exo-type alginate lyase from *Agrobacterium fabrum*), an arginine is reported to act as a neutralizer of the acidic group (Ochiai et al., 2010). However, the PL15 enzyme, adopts the two-domain $(\alpha/\alpha)_{5,6}$ toroid antiparallel β-sandwich fold. The completely different fold of ulvan lyases (LOR107 and PLSV3936 - seven bladed β-propeller) and that of the PL15 enzymes and yet similarity in the active site arrangement, indicates a convergent evolution of these enzymes toward the same active site. As mentioned above, other PL families (PL11 and PL22) with β-propeller folds are evolved to utilize the metal assisted mechanism, with a Ca^{2+} (PL11) and Mn^{2+} (PL22) ion being the neutralizer. Thus PL24 (LOR107) and PL25 (PLSV3936) family ulvan lyases expand the range of PL folds that adopt the histidine-tyrosine mechanism of catalysis.

Comparison with other seven bladed β propeller enzymes

A large number of proteins with a similar seven-bladed β-propeller fold have been identified using DALI (Holm and Rosenström, 2010) and deconSTRUCT (Zhang et al., 2010) servers. This fold was observed frequently in enzymes as well as in protein domains that bind other proteins. However, a good structural fit is limited to a subset of blades for most of them. BACOVA_03493 from the gut bacteria *Bacteroides ovatus* (PDB code: 4IRT, Joint Center for Structural Genomics) shows the best structural fit that encompasses all blades. These two structures, PLSV3936 and 4IRT, can be superimposed with an RMSD of 1.64 Å for 267 Cα atoms with all the blades showing good overlap **(Figure 7.3A)**. The structure-based alignment shows only 15% sequence identity, indicating a distant evolutionary relationship.

The structural comparison of PLSV3936 and BACOVA_03493 revealed that the highly conserved amino acids forming the active site of PLSV3936 are also present in BACOVA_03493. The key residues of PLSV3936 are indeed conserved, namely His90 (His123 in PLSV3936), His110 (His143), Tyr143 (Tyr188), Arg159 (Arg204), Tyr197 (Tyr246) **(Figure 7.3 F)**. Several other residues conserved in PL25 enzymes (PLSV3936), in addition to the active site, were also present in BACOVA_03493. Therefore, despite a low sequence identity between these two proteins, we rationalized that their catalytic mechanisms might be the same. This hypothesis is further supported by the analysis of the surrounding genes. The gene encoding BACOVA_03493, is located in the polysaccharide utilization loci and flanked by the glycoside hydrolase (GH) family enzymes namely; BACOVA_03492, a GH43 and BACOVA_03494, a GH105 family glycoside hydrolases. The GH105 members are known to hydrolyze the disaccharide end products of lyases. The purified BACOVA_03493 protein sample was obtained from The Scripps Research Institute (TSRI) and was tested for enzyme activity against the ulvan polysaccharide. However, no activity was detected, indicating that this enzyme is not an ulvan lyase. Enzyme activity of BACOVA_03493 was tested against heparin, chondroitin, hyaluronan and xanthan at the CERMAV (Centre de recherches sur les macromolécules végétales) and no activity was detected against these substrates. Unexpectedly, a glycoside hydrolase activity was detected for this enzyme and the active site was found on the opposite end (posterior side) of the β-propeller **(Figure 7.4C)** (Munoz-Munoz et al., 2017).

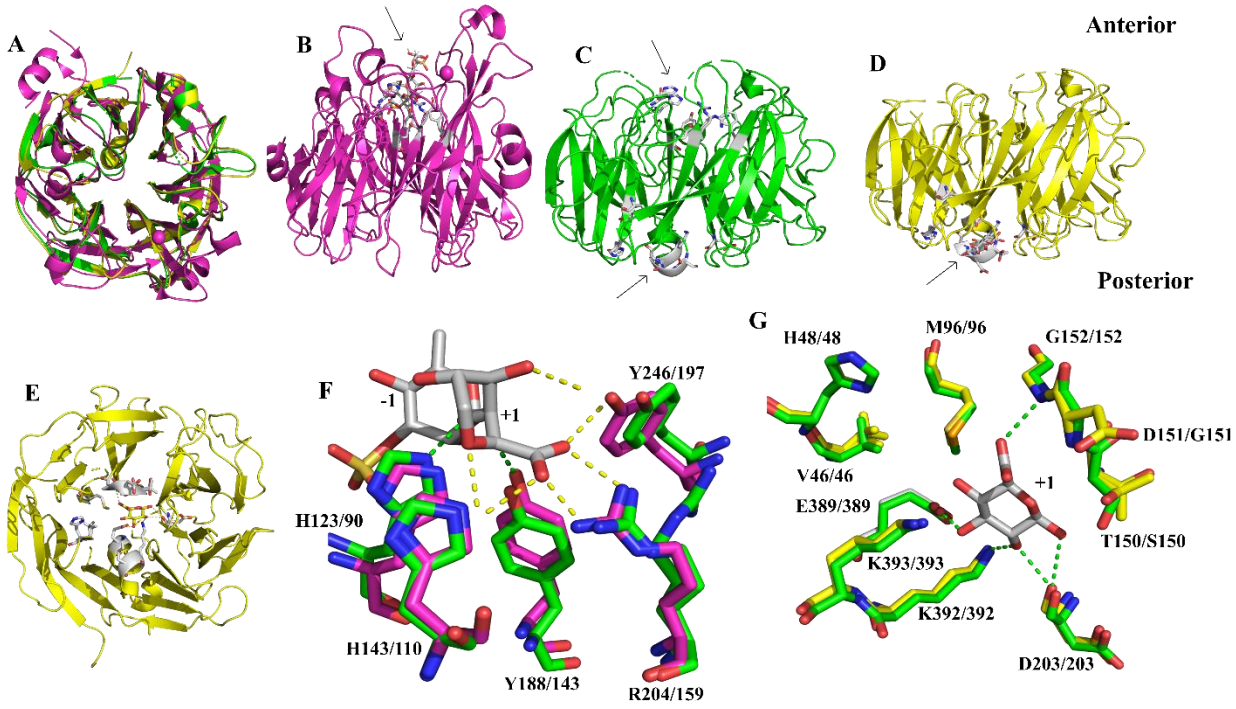

Figure 7:3 Structure comparison of PLSV3936 and hydrolase

(A) Cartoon representation of PLSV3936 (colored magenta), 4IRT (colored green) and 5MUL (colored green) superposition showing the propeller. **(B)** The active site of PLSV3936 containing the substrate at anterior side of propeller, indicating by arrow. The catalytic residue and the substrate are shown as sticks and colored white. **(C)** The active site of 4IRT and the conserved residues at anterior and posterior side of propeller, indicating by arrow. **(D)** The active site of 5MUL at posterior side of propeller. **(E)** Posterior side of 5MUL propeller and the catalytic site. The GlcA is colored yellow and the interacting residues are colored white. **(F)** Active site comparison showing the residue conservation of PLSV 3936 active site and anterior side of 4IRT. **(G)** Active site comparison showing the residue conservation of 5MUL active site and posterior side of 4IRT.

Interestingly, the hydrolase active site does not contain the canonical hydrolase catalytic apparatus (with two carboxylic groups in the active site) similar to BT3686 (PDB code: 5MUL), a member of GH145 family **(Figure 7.3D)**. This is quite unusual and suggests an evolutionary switch of the active site and catalytic mechanism in a fold that was performing a different function.

8. Summary and future perspectives

The work presented in this thesis involves the structural studies on four PL family enzymes. The PL enzymes degrade the uronic acid containing polysaccharide through a β-elimination mechanism. The general steps involved in this lytic degradation are: (1) neutralization of the acidic group (2) abstraction of a C5 proton and (3) transfer of a proton to the leaving group

In the first part of this work, the crystal structure of HepIII (a GAG degrading PL12 family enzyme) has been determined. HepIII adopts the two domain $(\alpha/\alpha)_n$ toroid+β-sandwich fold. A structural comparison with homologs showed that the His-Tyr catalytic dyad is conserved in HepIII. The unsuccessful attempts to obtain the HepIII-substrate complex structure was explained by the conformational flexibility seen in this enzyme. The results of our NMA calculation on the three HepIII structures identified a large-scale domain movement; with the hinge region near Asn281. In addition, the NMA predicts the possible active conformation of these lyases and this conformational change could be induced by the binding of longer GAG substrates.

In the second part of this work, the first crystal structure of three ulvan lyase enzymes: LOR107, PLSV3936 and NLR48 from marine bacteria have been determined. All three enzymes share low sequence identity and degrade the marine polysaccharide, ulvan. Our structure data show that LOR107 and PLSV3936 adopt the β-propeller fold and NLR48 adopts the β-jelly roll fold. However, the His-Tyr catalytic dyad is conserved in all these enzymes.

The complex structures of LOR107 with substrate suggest that Arg259 plays the role of neutralizer and His146 can play the dual base/acid role in the lytic degradation of the ulvan substrate. His167, Tyr243 are the other important residues facilitating the catalytic mechanism. A complex structure of PLSV3936 with substrate suggests that Arg 204 is the neutralizer of the carboxylic group, Tyr188 and His123 plays the catalytic base/acid role, respectively. However, the structural data in NLR48 suggests that the possible neutralizer of the carboxylic group is a Gln137. Tyr258 act as a base and acid for the cleavage near GlcA, while for IdoA the base is Lys139 and the acid is Tyr258.

Overall the structural data suggests that the His/Tyr catalytic machinery is conserved among ulvan lyase enzymes. Differences seen in the active site reflect the plasticity adopted by

these enzymes to accommodate the heterogeneity of ulvan. Our structural data will serve as a template and aid researchers in engineering ulvan lyases to make them more suitable for industrial applications.

Table 8-1 Summary of research contribution

PL	Family	Substrate	Fold	Mechanism	PDB	Neutralizer	Base	Acid
12	Heparin lyase III	Heparan sulfate	$(\alpha/\alpha)_6$ toroid + anti-parallel β-sheet	His/Tyr	5JMD	Asn247	Tyr303	
24	Ulvan lyase	Ulvan	β-propeller (7 blades)	His/Tyr	6BYP 6BYX 6BYT	Arg259	His146	
25	Ulvan lyase	Ulvan	β-propeller (7 blades)	His/Tyr	5UAM 5UAS	Arg204	Tyr188	His123
28	Ulvan lyase	Ulvan	β-jelly roll	His/Tyr	6D2C 6D3U	Gln137	Lys139 Tyr258	Tyr258

9. References

Abbott, D.W., Boraston, A.B. (2008). Structural Biology of Pectin Degradation by Enterobacteriaceae. Microbiol. Mol. Biol. Rev. *72*, 301–316.

Abbott, D.W., Boraston, A.B. (2007). A family 2 pectate lyase displays a rare fold and transition metal-assisted β-elimination. J. Biol. Chem. *282*, 35328–35336.

Abbott, D.W., Gilbert, H.J., Boraston, A.B. (2010). The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate β-elimination. J. Biol. Chem. *285*, 39029–39038.

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. (2011). The Phenix software for automated determination of macromolecular structures. Methods *55*, 94–106.

Ahmad, N.H., Mustafa, S., Man, Y.B.C. (2015). Microbial polysaccharides and their modification approaches: A review. Int. J. Food Prop. *18*, 332–347.

Akita, M., Suzuki, A., Kobayashi, T., Ito, S., Yamane, T. (2001). The first structure of pectate lyase belonging to polysaccharide lyase family 3. Acta Crystallogr. Sect. D Biol. Crystallogr. *57*, 1786–1792.

Alkawash, M.A., Soothill, J.S., Schiller, N.L. (2006). Alginate lyase enhances antibiotic killing of mucoid Pseudomonas aeruginosa in biofilms. Apmis *114*, 131–138.

Alves, A., Sousa, R.A., Reis, R.L. (2013). Processing of degradable ulvan 3D porous structures for biomedical applications. J. Biomed. Mater. Res. - Part A *101* A, 998–1006.

Bahar, I., Lezon, T.R., Yang, L.-W., Eyal, E. (2010). Global Dynamics of Proteins: Bridging Between Structure and Function. Annu. Rev. Biophys. Biomol. Struct. *9*, 23–42.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc. Natl. Acad. Sci. *98*, 10037–10041.

Bell, M.R., Engleka, M.J., Malik, A., Strickler, J.E. (2013). To fuse or not to fuse: What is your purpose? Protein Sci. *22*, 1466-1477

Bertozzi, C.R., Kiessling, L.L. (2001). Chemical glycobiology. Science. *29*, 2357-2364.

Bikker, P., van Krimpen, M.M., van Wikselaar, P., Houweling-Tan, B., Scaccia, N., van Hal, J.W., Huijgen, W.J.J., Cone, J.W., López-Contreras, A.M. (2016). Biorefinery of the green seaweed Ulva lactuca to produce animal feed, chemicals and biofuels. J. Appl. Phycol. *28*, 3511–3525.

Brünger, A.T. (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. Nature. *355*, 472–475.

Cartmell, A., Lowe, E.C., Baslé, A., Firbank, S.J., Ndeh, D.A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J.E., Czjzek, M., Gilbert, H.J., Bolam, D.N. (2017). How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. Proc. Natl. Acad. Sci. *114*, 7037–7042.

Charnock, S.J., Brown, I.E., Turkenburg, J.P., Black, G.W., Davies, G.J. (2002). Convergent evolution sheds light on the anti-beta -elimination mechanism common to family 1 and 10 polysaccharide lyases. Proc. Natl. Acad. Sci. U. S. A. *99,* 12067–12072.

Chayen, N.E. (2004). Turning protein crystallisation from an art into a science. Curr. Opin. Struct. Biol. *14*, 577–83.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 12–21.

Colle, P.N., Sassi, J.F., Rogniaux, H., Marfaing, H., Helbert, W. (2011). Ulvan lyases isolated from the flavobacteria Persicivirga ulvanivorans are the first members of a new polysaccharide lyase family. J. Biol. Chem. *286*, 42063–42071.

Collén, P.N., Jeudy, A., Sassi, J.F., Groisillier, A., Czjzek, M., Coutinho, P.M., Helbert, W. (2014). A novel unsaturated β-glucuronyl hydrolase involved in ulvan degradation unveils the versatility of stereochemistry requirements in family GH105. J. Biol. Chem. *289*, 6199–6211.

Corbett, D., Roberts, I.S. (2009). The role of microbial polysaccharides in host-pathogen interaction. F1000 Biol. Rep. *1*, 30.

Creze, C., Castang, S., Derivery, E., Haser, R., Hugouvieux-Cotte-Pattat, N., Shevchik, V.E., Gouet, P. (2008). The crystal structure of pectate lyase peli from soft rot pathogen Erwinia chrysanthemi in complex with its substrate. J. Biol. Chem. *283*, 18260–18268.

Cummings, J.H., Stephen, A.M. (2007). Carbohydrate terminology and classification. Eur. J. Clin. Nutr. *61*, S5–S18.

Cummings, R.D., Pierce, J.M. (2010). Handbook of Glycomics.

Cunha, L., Grenha, A. (2016). Sulfated seaweed polysaccharides as multifunctional materials in drug delivery applications. Mar. Drugs.*14*, 42

Dauter, Z. (2017). Collection of X-ray diffraction data from macromolecular crystals, Protein crystallography: Methods and protocol, 165–184.

Davies, G.J., Wilson, K.S., Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. *321*, 557–559.

De Armas, H.N., Verboven, C., De Ranter, C., Desair, J., Vande Broek, A., Vanderleyden, J., Rabijns, A. (2004). Azospirillum irakense pectate lyase displays a toroidal fold. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 999–1007.

Delattre, C., Michaud, P., Keller, C., Elboutachfaiti, R., Beven, L., Courtois, B., Courtois, J. (2006). Purification and characterization of a novel glucuronan lyase from Trichoderma sp. GL2. Appl. Microbiol. Biotechnol. *70*, 437–443.

Desai, U.R., Wang, H.M., Linhardt, R.J. (1993). Substrate specificity of the heparin lyases from Flavobacterium heparinum. Arch. Biochem. Biophys. *306*, 461–8.

Dong, S., Wei, T. Di, Chen, X.L., Li, C.Y., Wang, P., Xie, B. Bin, Qin, Q.L., Zhang, X.Y., Pang, X.H., Zhou, B.C., Zhang, Y.Z. (2014). Molecular insight into the role of the N-terminal extension

in the maturation, substrate recognition, and catalysis of a bacterial alginate lyase from polysaccharide lyase family 18. J. Biol. Chem. *289*, 29558–29569.

Dong, W., Lu, W., Mckeehan, W.L., Luo, Y., Ye, S. (2012). Structural basis of heparan sulfate-specific degradation by heparinase III. Protein & cell *12*, 950-961.

Dumas, B., Jaulneau, V., Lafitte, C., Jacquet, C., Fournier, S., Salamagne, S., Briand, X., Esquerré-Tugayé, M.T. (2010). Ulvan, a sulfated polysaccharide from green algae, activates plant immunity through the jasmonic acid signaling pathway. J. Biomed. Biotechnol.

Dzvova, N., Colmer-Hamood, J.A., Griswold, J.A., Hamood, A.N. (2017). Isolation and characterization of HepP: A virulence-related Pseudomonas aeruginosa heparinase. BMC Microbiol. *17*, 1–16.

Dzvova, N., Colmer-Hamood, J.A., Griswold, J.A., Hamooda, A.N. (2018). Heparinase is essential for Pseudomonas aeruginosa virulence during thermal injury and infection. Infect. Immun. *86*, 1–13.

Emsley, P., Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 486–501.

Eschenfeldt, W.H., Lucy, S., Millard, C.S., Joachimiak, A., Mark, I.D. (2009). A family of LIC vectors for high-throughput cloning and purification of proteins. Methods Mol. Biol. *498*, 105–115.

Evans, P., McCoy, A. (2008). An introduction to molecular replacement. Acta Crystallogr. D. Biol. Crystallogr. *64*, 1–10.

Evans, P.R., Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. *69*, 1204–1214.

Feingold, D.S., Bentley, R. (1987). Conformational aspects of the reaction mechanisms of polysaccharide lyases and epimerases. FEBS Lett. *223*, 207–211.

Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. Gut microbes. *4*, 289-306.

Fodje, M., Grochulski, P., Janzen, K., Labiuk, S., Gorin, J., Berg, R. (2014). 08B1-1: An automated beamline for macromolecular crystallography experiments at the Canadian Light Source. J. Synchrotron Radiat. *21*, 633-637.

Foran, E., Buravenkov, V., Kopel, M., Mizrahi, N., Shoshani, S., Helbert, W., Banin, E. (2017). Functional characterization of a novel “ulvan utilization loci” found in Alteromonas sp. LOR genome. Algal Res. *25*, 39–46.

Funderburgh, J.L. (2002). Keratan sulfate biosynthesis. IUBMB Life *54*, 187–194.

Gacesa, P. (1987). Alginate-modifying enzymes. A proposed unified mechanism of action for the lyases and epimerases. FEBS Lett. *212*, 199–202.

Gandhi, N.S., Mancera, R.L. (2008). The structure of glycosaminoglycans and their interactions with proteins. Chem. Biol. Drug Des. *72*, 455–82.

García, B., Merayo-Lloves, J., Martin, C., Alcalde, I., Quirós, L.M., Vazquez, F. (2016). Surface proteoglycans as mediators in bacterial pathogens infections. Front. Microbiol. *7*, 220.

Garman, E.F. (2014). Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. Science *343*, 1102–1108.

Garron, M.-L., Cygler, M. (2014). Uronic polysaccharide degrading enzymes. Curr. Opin. Struct. Biol. *28*, 87–95.

Garron, M.L., Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology *20*, 1547–1573.

Gilbert, H.J. (2010). The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. *153*, 444–455.

Grochulski, P., Fodje, M.N., Gorin, J., Labiuk, S.L., Berg, R. (2011). Beamline 08ID-1, the prime beamline of the Canadian macromolecular crystallography facility. J. Synchrotron Radiat. *18*, 681–684.

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W., Brumer, H. (2017). Polysaccharide Utilization Loci: Fuelling microbial communities. J. Bacteriol. *199*, 00860-16.

Han, Y.-H., Garron, M.-L., Kim, H.-Y., Kim, W.-S., Zhang, Z., Ryu, K.-S., Shaya, D., Xiao, Z., Cheong, C., Kim, Y.S., Linhardt, R.J., Jeon, Y.H., Cygler, M. (2009). Structural snapshots of heparin depolymerization by heparin lyase I. J. Biol. Chem. *284*, 34019–27.

Hassan, S., Shevchik, V.E., Robert, X., Hugouvieux-Cotte-Pattat, N. (2013). PelN is a new pectate lyase of dickeya dadantii with unusual characteristics. J. Bacteriol. *195*, 2197–2206.

Hassell, A.M., An, G., Bledsoe, R.K., Bynum, J.M., Carter, H.L., Deng, S.J.J., Gampe, R.T., Grisard, T.E., Madauss, K.P., Nolte, R.T., Rocque, W.J., Wang, L., Weaver, K.L., Williams, S.P., Wisely, G.B., Xu, R., Shewchuk, L.M. (2006). Crystallization of protein-ligand complexes, Acta Crystallographica Section D: Biological Crystallography. *63*, 72–79.

Hehemann, J.H., Boraston, A.B., Czjzek, M. (2014). A sweet new wave: Structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr. Opin. Struct. Biol. *28*, 77-86.

Hemsworth, G.R., Dejean, G., Davies, G.J., Brumer, H., 2016. Learning from microbial strategies for polysaccharide degradation. Biochem. Soc. Trans. *44*, 94–108.

Hendrickson, W. a, Horton, J.R., LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. *9*, 1665–1672.

Herron, S.R., Benen, J. a, Scavetta, R.D., Visser, J., Jurnak, F. (2000). Structure and function of pectic enzymes: virulence factors of plant pathogens. Proc. Natl. Acad. Sci. U. S. A. *97*, 8762–8769.

Holm, L., Rosenström, P. (2010). Dali server: Conservation mapping in 3D. Nucleic Acids Res.

38, 545-549.

Huang, W., Boju, L., Tkalec, L., Su, H., Yang, H.O., Gunay, N.S., Linhardt, R.J., Yeong, S.K., Matte, A., Cygler, M. (2001). Active site of chondroitin AC lyase revealed by the structure of enzyme oligosaccharide complexes and mutagenesis. Biochemistry *40*, 2359–2372.

Huang, W., Lunin, V. V., Li, Y., Suzuki, S., Sugiura, N., Miyazono, H., Cygler, M. (2003). Crystal structure of Proteus vulgaris chondroitin sulfate ABC lyase I at 1.9 Å resolution. J. Mol. Biol. *328*, 623–634.

Huang, W., Matte, A., Li, Y., Kim, Y.S., Linhardt, R.J., Su, H., Cygler, M. (1999). Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 Å resolution. J. Mol. Biol. *294*, 1257–1269.

Huegel, J., Sgariglia, F., Enomoto-Iwamoto, M., Koyama, E., Dormans, J.P., Pacifici, M. (2013). Heparan sulfate in skeletal development, growth, and pathology: the case of hereditary multiple exostoses. Dev. Dyn. *242*, 1021–32.

Hugouvieux-Cotte-Pattat, N., Condemine, G., Shevchik, V.E. (2014). Bacterial pectate lyases, structural and functional diversity. Environ. Microbiol. Rep. *6*, 427–440.

Jain, S., Ohman, D.E. (2005). Role of an alginate lyase for alginate transport in mucoid Pseudomonas aeruginosa. Infect. Immun. *73*, 6429–6436.

Jenkins, J., Shevchik, V.E., Hugouvieux-Cotte-Pattat, N., Pickersgill, R.W. (2004). The Crystal Structure of Pectate Lyase Pel9A from Erwinia chrysanthemi. J. Biol. Chem. *279*, 9139–9145.

Jensen, M.H., Otten, H., Christensen, U., Borchert, T. V., Christensen, L.L.H., Larsen, S., Leggio, L. Lo. (2010). Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from aspergillus aculeatus. J. Mol. Biol. *404*, 100–111.

Jinno, A., Park, P.W. (2015). Role of glycosaminoglycans in infectious disease. Methods Mol. Biol. *1229*, 567–585.

Jongkees, S.A.K., Withers, S.G. (2011). Glycoside cleavage by a new mechanism in unsaturated glucuronyl hydrolases. J. Am. Chem. Soc. *133*, 19334-19337.

Kabsch, W. (2010). Xds. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 125–132.

Karplus, P.A., Diederichs, K. (2012). Linking crystallographic model and data quality. Science. *336*, 1030–1033.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature. *474*, 327-336

Kawaguchi, Y., Sugiura, N., Kimata, K., Kimura, M., Kakuta, Y. (2013). The crystal structure of novel chondroitin lyase ODV-E66, a baculovirus envelope protein. FEBS Lett. *587*, 3943-3948.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. *181*, 662–666.

Konno, N., Ishida, T., Igarashi, K., Fushinobu, S., Habu, N., Samejima, M., Isogai, A. (2009). Crystal structure of polysaccharide lyase family 20 endo-β-1,4-glucuronan lyase from the

filamentous fungus Trichoderma reesei. FEBS Lett. *583*, 1323–1326.

Kopel, M., Helbert, W., Belnik, Y., Buravenkov, V., Herman, A., Banin, E. (2016). New Family of Ulvan Lyases Identified in Three Isolates from the Alteromonadales Order. J. Biol. Chem. *291*, 5871-5878.

Kopel, M., Helbert, W., Henrissat, B., Doniger, T., Banin, E. (2014a). Draft Genome Sequence of Nonlabens ulvanivorans, an Ulvan-Degrading Bacterium. Genome Announc. *2*, 3–4.

Kopel, M., Helbert, W., Henrissat, B., Doniger, T., Banin, E. (2014b). Draft genome sequence of *Pseudoalteromonas* strain PLSV, an ulvan-degrading bacterium. Genome Announc. *2*, 6–7.

Kopel, M., Helbert, W., Henrissat, B., Doniger, T., Banin, E. (2014c). Draft genome sequences of two ulvan-degrading isolates, strains LTR and LOR , that belong to the *Alteromonas* genus. Genome Announc. *2*, 13–15.

Koropatkin, N.M., Cameron, E.A., Martens, E.C. (2012). How glycan metabolism shapes the human gut microbiota. Nat. Rev. Microbiol.*10,* 323-335

Krojer, T., Pike, A.C.W., Von Delft, F. (2013). Squeezing the most from every crystal: The fine details of data collection. Acta Crystallogr. Sect. D Biol. Crystallogr. *69,* 1303–1313.

Lahaye, M., Brunel, M., Bonnin, E. (1997). Fine chemical structure analysis of oligosaccharides produced by an ulvan-lyase degradation of the water-soluble cell-wall polysaccharides from Ulua sp. (Ulvales, Chlorophyta). Carbohydr. Res. *304,* 325–333.

Lahaye, M., Robic, A. (2007). Structure and function properties of Ulvan, a polysaccharide from green seaweeds. Biomacromolecules *8,* 1765–1774.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., Ben-Tal, N. (2005). ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. *33*, W299-W302.

Langer, G., Cohen, S.X., Lamzin, V.S., Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. *3,* 1171–1179.

Laskowski, R.A., Hutchinson, E.G., Michie, A.D., Wallace, A.C., Jones, M.L., Thornton, J.M. (1997). PDBsum: A Web-based database of summaries and analyses of all PDB structures. Trends Biochem. Sci. *22,* 488-490.

Laurienzo, P. (2010). Marine polysaccharides in pharmaceutical applications: An overview. Mar. Drugs. *8,* 2435-2465.

Lebedev, A.A., Young, P., Isupov, M.N., Moroz, O. V., Vagin, A.A., Murshudov, G.N. (2012). JLigand: A graphical tool for the CCP4 template-restraint library. Acta Crystallogr. Sect. D Biol. Crystallogr. *68,* 431–440.

Lebendiker, M., Danieli, T. (2014). Production of prone-to-aggregate proteins. FEBS Lett. *588,* 236-246

Leid, J.G., Willson, C.J., Shirtliff, M.E., Hassett, D.J., Parsek, M.R., Jeffers, A.K. (2005). The Exopolysaccharide Alginate Protects Pseudomonas aeruginosa Biofilm Bacteria from IFN- -Mediated Macrophage Killing. J. Immunol. *175,* 7512–7518.

Li, S., Kelly, S.J., Lamani, E., Ferraroni, M., Jedrzejas, M.J. (2000). Structural basis of hyaluronan degradation by Streptococcus pneumoniae hyaluronate lyase. EMBO J. *19*, 1228–40.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. *42*, 490-495.

Lunin, V. V., Li, Y., Linhardt, R.J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A.W., Cygler, M. (2004). High-resolution crystal structure of Arthrobacter aurescens chondroitin AC lyase: An enzyme-substrate complex defines the catalytic mechanism. J. Mol. Biol. *337*, 367–386.

MacDonald, L.C., Berger, B.W. (2014). A polysaccharide lyase from Stenotrophomonas maltophilia with a unique, pH-regulated substrate specificity. J. Biol. Chem. *289*, 312–325.

Makhoba, X.H., Mthembu, M.S. (2015). Molecular Chaperone Assisted Expression Systems: Obtaining Pure Soluble and Active Recombinant Proteins for Structural and Therapeutic Purposes. J. Proteomics Bioinform. *08*, 212–216.

Manivasagan, P., Oh, J. (2016). Marine polysaccharide-based nanomaterials as a novel source of nanobiotechnological applications. Int. J. Biol. Macromol. *82*, 315–327.

Marion, C., Stewart, J.M., Tazi, M.F., Burnaugh, A.M., Linke, C.M., Woodiga, S.A., King, S.J. (2012). Streptococcus pneumoniae can utilize multiple sources of hyaluronic acid for growth. Infect. Immun. *80*, 1390–1398.

Matard-Mann, M., Bernard, T., Leroux, C., Barbeyron, T., Larocque, R., Préchoux, A., Jeudy, A., Jam, M., Nyvall-Collén, P., Michel, G., Czjzek, M. (2017). Structural insights into marine carbohydrate degradation by family GH16 kappa-carrageenases. J. Biol. Chem. *292*, 19919-19934.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McDonough, M.A., Kadirvelraj, R., Harris, P., Poulsen, J.C.N., Larsen, S. (2004). Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. FEBS Lett. *565*, 188–194.

McLean, R., Hobbs, J.K., Suits, M.D., Tuomivaara, S.T., Jones, D.R., Boraston, A.B., Abbott, D.W. (2015). Functional analyses of resurrected and contemporary enzymes illuminate an evolutionary path for the emergence of exolysis in polysaccharide lyase family. J. Biol. Chem. *290*, 21231–21243.

Medipally, S.R., Yusoff, F.M., Banerjee, S., Shariff, M. (2015). Microalgae as sustainable renewable energy feedstock for biofuel production. Biomed Res. Int.

Michel, G., Pojasek, K., Li, Y., Sulea, T., Linhardt, R.J., Raman, R., Prabhakar, V., Sasisekharan, R., Cygler, M. (2004). The structure of chondroitin B lyase complexed with glycosaminoglycan oligosaccharides unravels a calcium-dependent catalytic machinery. J. Biol. Chem. *279*, 32882–32896.

Mikami, T., Kitagawa, H. (2013). Biosynthesis and function of chondroitin sulfate. Biochim. Biophys. Acta - Gen. Subj.

Minor, W., Cymborowski, M., Otwinowski, Z., Chruszcz, M. (2006). HKL-3000: The integration

of data reduction and structure solution - From diffraction images to an initial model in minutes. Acta Crystallogr. Sect. D Biol. Crystallogr. *62*, 859–866.

Mishra, P., Prem Kumar, R., Ethayathulla, A.S., Singh, N., Sharma, S., Perbandt, M., Betzel, C., Kaur, P., Srinivasan, A., Bhakuni, V., Singh, T.P. (2009). Polysaccharide binding sites in hyaluronate lyase - Crystal structures of native phage-encoded hyaluronate lyase and its complexes with ascorbic acid and lactose. FEBS J. *276*, 3392–3402.

Mizumoto, S., Sugahara, K. (2012). Glycosaminoglycan chain analysis and characterization (glycosylation/epimerization). Methods Mol. Biol. *836*, 99–115.

Mohnen, D. (2008). Pectin structure and biosynthesis. Curr. Opin. Plant Biol. *11*, 266–277.

Morelli, A., Betti, M., Puppi, D., Chiellini, F. (2016). Design, preparation and characterization of ulvan based thermosensitive hydrogels. Carbohydr. Polym. *136*, 1108–1117.

Moriarty, N.W., Grosse-Kunstleve, R.W., Adams, P.D. (2009). Electronic ligand builder and optimization workbench (eLBOW): A tool for ligand coordinate and restraint generation. Acta Crystallogr. Sect. D Biol. Crystallogr. *65*, 1074–1080.

Morris, G., Harding, S. (2009). Polysaccharides, Microbial. In Encyclopedia of Microbiology. Schaechter, ed., 482–494.

Müller, I. (2017). Guidelines for the successful generation of protein-ligand complex crystals. Acta Crystallogr. Sect. D Struct. Biol. *73*, 79-92.

Munoz-Munoz, J., Cartmell, A., Terrapon, N., Henrissat, B., Gilbert, H.J. (2017). Unusual active site location and catalytic apparatus in a glycoside hydrolase family. Proc. Natl. Acad. Sci. *114*, 4936–4941.

Nalamothu, N., Potluri, A., Muppalla, M.B. (2014). Review on Marine Alginates and Its Applications. Indo Am. J. Pharm. Res. *4*, 4006–4015.

Ndeh, D., Gilbert, H.J. (2018). Biochemistry of Complex Glycan Depolymerisation by the Human Gut Microbiota. FEMS Microbiol. Rev. *42*, 146–164.

Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R.A., Zhu, Y., O'Neill, M.A., Urbanowicz, B.R., York, W.S., Davies, G.J., Abbott, D.W., Ralet, M.C., Martens, E.C., Henrissat, B., Gilbert, H.J. (2017). Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature *544*, 65–70.

Ochiai, A., Itoh, T., Maruyama, Y., Kawamata, A., Mikami, B., Hashimoto, W., Murata, K. (2007). A novel structural fold in polysaccharide lyases: Bacillus subtilis family 11 rhamnogalacturonan lyase YesW with an eight-bladed β-propeller. J. Biol. Chem. *282*, 37134–37145.

Ochiai, A., Itoh, T., Mikami, B., Hashimoto, W., Murata, K. (2009). Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases. J. Biol. Chem. *284*, 10181–10189.

Ochiai, A., Yamasaki, M., Mikami, B., Hashimoto, W., Murata, K. (2010). Crystal structure of exotype alginate lyase Atu3025 from Agrobacterium tumefaciens. J. Biol. Chem. *285*, 24519–

24528.

Ogura, K., Yamasaki, M., Mikami, B., Hashimoto, W., Murata, K. (2008). Substrate recognition by family 7 alginate lyase from Sphingomonas sp. A1. J. Mol. Biol. *380*, 373–385.

Ogura, K., Yamasaki, M., Yamada, T., Mikami, B., Hashimoto, W., Murata, K. (2009). Crystal structure of family 14 polysaccharide lyase with pH-dependent modes of action. J. Biol. Chem. *284*, 35572–35579.

Osawa, T., Matsubara, Y., Muramatsu, T., Kimura, M., Kakuta, Y. (2005). Crystal structure of the alginate (Poly α-L-guluronate) lyase from Corynebacterium sp. at 1.2 Å resolution. J. Mol. Biol. *345*, 1111–1118.

Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307-26

Park, D., Jagtap, S., Nair, S.K. (2014). Structure of a PL17 family alginate lyase demonstrates functional similarities among exotype depolymerases. J. Biol. Chem. *289*, 8645–8655.

Pflugrath, J.W. (2015). Practical macromolecular cryocrystallography. Acta Crystallogr. Sect. FStructural Biol. Commun. *71*, 622–642.

Powell, H.R. (2017). X-ray data processing. Biosci. Rep. *37*, BSR20170227.

Pritt, B., O'Brien, L., Winn, W. (2007). Mucoid Pseudomonas in cystic fibrosis. Am. J. Clin. Pathol. *128*, 32–34.

Proteau, A., Shi, R., Cygler, M. (2010). Application of dynamic light scattering in protein crystallization. Curr. Protoc. Protein Sci. *61*, 17.10.1-17.10.9.

Rani, A., Patel, S., Goyal, A. (2017). Chondroitin sulfate (CS) lyases: Structure, function and application in therapeutics. Curr. Protein Pept. Sci. *19*, 22-33.

Ridley, B.L., O'Neill, M.A., Mohnen, D. (2001). Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling, Phytochemistry. *57*, 929-967.

Rigden, D.J., Jedrzejas, M.J. (2003). Structures of streptococcus pneumoniae hyaluronate lyase in complex with chondroitin and chondroitin sulfate disaccharides: Insights into specificity and mechanism of action. J. Biol. Chem. *278*, 50596–50606.

Rigden, D.J., Littlejohn, J.E., Joshi, H. V., de Groot, B.L., Jedrzejas, M.J. (2006). Alternate Structural Conformations of Streptococcus pneumoniae Hyaluronan Lyase: Insights into Enzyme Flexibility and Underlying Molecular Mechanism of Action. J. Mol. Biol. *358*, 1165–1178.

Rodgers, D.W. (1994). Cryocrystallography. Structure *2*, 1135–1140.

Rodrigues, M.L., Nimrichter, L., Cordero, R.J.B., Casadevall, A. (2011). Fungal polysaccharides: Biological activity beyond the usual structural properties. Front. Microbiol. *2*, 1–4.

Rosano, G.L., Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Front. Microbiol. *5*,172

Rupp, B. (2015). Origin and use of crystallization phase diagrams. Acta Crystallogr. Sect. FStructural Biol. Commun. *71*, 247–260.

Rye, C.S., Matte, A., Cygler, M., Withers, S.G. (2006). An atypical approach identifies TYR234 as the key base catalyst in chondroitin AC lyase. ChemBioChem *7*, 631–637.

Rye, C.S., Withers, S.G. (2002). Elucidation of the Mechanism of Polysaccharide Cleavage by Chondroitin AC Lyase from *Flavobacterium h eparinum*. J. Am. Chem. Soc. *124*, 9756–9767.

Rye, C.S., Withers, S.G. (2000). Glycosidase mechanisms. Current opinion in chemical biology, *4*, 573-580.

Sarrazin, S., Lamanna, W.C., Esko, J.D. (2011). Heparan sulfate proteoglycans. Cold Spring Harb. Perspect. Biol. *3*, 1–33.

Sava, I.G., Zhang, F., Toma, I., Theilacker, C., Li, B., Baumert, T.F., Holst, O., Linhardt, R.J., Huebner, J. (2009). Novel interactions of glycosaminoglycans and bacterial glycolipids mediate binding of enterococci to human cells. J. Biol. Chem. *284*, 18194–18201.

Schwartz, N.B., Domowicz, M. (2002). Chondrodysplasias due to proteoglycan defects. Glycobiology *12*, 57R–68R.

Senni, K., Pereira, J., Gueniche, F., Delbarre-Ladrat, C., Sinquin, C., Ratiskol, J., Godeau, G., Fischer, A.M., Helley, D., Colliec-Jouault, S. (2011). Marine polysaccharides: A source of bioactive molecules for cell therapy and tissue engineering. Mar. Drugs. *9*, 1664-1681.

Seyedarabi, A., To, T.T., Ali, S., Hussain, S., Fries, M., Madsen, R., Clausen, M.H., Teixteira, S., Brocklehurst, K., Pickersgill, R.W. (2010). Structural insights into substrate specificity and the anti β-elimination mechanism of pectate lyase. Biochemistry *49*, 539–546.

Shaw Stewart, P.D., Kolek, S.A., Briggs, R.A., Chayen, N.E., Baldock, P.F.M. (2011). Random microseeding: A theoretical and practical exploration of seed stability and seeding techniques for successful protein crystallization. Cryst. Growth Des. *11*, 3432–3441.

Shaya, D., Hahn, B.S., Bjerkan, T.M., Kim, W.S., Park, N.Y., Sim, J.S., Kim, Y.S., Cygler, M. (2008). Composite active site of chondroitin lyase ABC accepting both epimers of uronic acid. Glycobiology *18*, 270–277.

Shaya, D., Tocilj, A., Li, Y., Myette, J., Venkataraman, G., Sasisekharan, R., Cygler, M. (2006). Crystal structure of heparinase II from Pedobacter heparinus and its complex with a disaccharide product. J. Biol. Chem. *281*, 15525–15535.

Shaya, D., Zhao, W., Garron, M., Xiao, Z., Cui, Q., Zhang, Z., Sulea, T., Linhardt, R.J., Cygler, M. (2010). Catalytic Mechanism of Heparinase II Investigated by Site-directed Mutagenesis and the Crystal Structure with Its Substrate. J. Biol. Chem. *285*, 20051–20061.

Shriver, Z., Capila, I., Venkataraman, G., Sasisekharan, R. (2012). Heparin and Heparan sulfate: Analyzing structure and microheterogeneity. Handb. Exp. Pharmacol. *207*, 159–176.

Singh, A., Olsen, S.I. (2011). A critical review of biochemical conversion, sustainability and life cycle assessment of algal biofuels. Appl. Energy *88*, 3548–3555.

Skubák, P., Pannu, N.S. (2013). Automatic protein structure solution from weak X-ray data. Nat. Commun. *4*, 2777.

Sledz, P., Zheng, H., Murzyn, K., Chruszcz, M., Zimmerman, M.D., Chordia, M.D., Joachimiak,

A., Minor, W. (2010). New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. Protein Sci. *19*, 1395–1404.

Smith, N.L., Taylor, E.J., Lindsay, A.-M., Charnock, S.J., Turkenburg, J.P., Dodson, E.J., Davies, G.J., Black, G.W. (2005). Structure of a group A streptococcal phage-encoded virulence factor reveals a catalytically active triple-stranded-helix. Proc. Natl. Acad. Sci. *102*, 17652–17657.

Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., Kitagawa, H. (2003). Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. Curr. Opin. Struct. Biol. *13*, 612–620.

Suhre, K., Sanejouand, Y. (2004). ElNemo: a normal mode web server for protein ElNe movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. *32*, 610–614.

Taylor, G. (2003). The phase problem, in: Acta Crystallographica - Section D Biological Crystallography. *59*, 1881–1890.

Taylor, G.L. (2010). Introduction to phasing. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 325–338.

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P. V., Zwart, P.H., Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. Acta Crystallogr. Sect. D Biol. Crystallogr. *65*, 582–601.

Thallinger, B., Prasetyo, E.N., Nyanhongo, G.S., Guebitz, G.M. (2013). Antimicrobial enzymes: An emerging strategy to fight microbes and microbial biofilms. Biotechnol. J. *8*, 97–109.

Thelin, M.A., Bartolini, B., Axelsson, J., Gustafsson, R., Tykesson, E., Pera, E., Oldberg, Å., MacCarana, M., Malmstrom, A. (2013). Biological functions of iduronic acid in chondroitin/dermatan sulfate. FEBS J. *280*, 2431–2446.

Thomas, F., Lundqvist, L.C.E., Jam, M., Jeudy, A., Barbeyron, T., Sandström, C., Michel, G., Czjzek, M. (2013). Comparative characterization of two marine alginate lyases from zobellia galactanivorans reveals distinct modes of action and exquisite adaptation to their natural substrate. J. Biol. Chem. *288*, 23021–23037.

Till, M., Robson, A., Byrne, M.J., Nair, A. V, Kolek, S.A., Shaw Stewart, P.D., Race, P.R. (2013). Improving the success rate of protein crystallization by random microseed matrix screening. J. Vis. Exp. 78.

Trowbridge, J.M., Gallo, R.L. (2002). Dermatan sulfate: new functions from an old glycosaminoglycan. Glycobiology *12*, 117R–25R.

Ulaganathan, T., Shi, R., Yao, D., Gu, R.X., Garron, M.L., Cherney, M., Tieleman, D.P., Sterner, E., Li, G., Li, L., Linhardt, R.J., Cygler, M. (2017). Conformational flexibility of PL12 family heparinases: Structure and substrate specificity of heparinase III from Bacteroides thetaiotaomicron (BT4657). Glycob. *27*, 176–187.

Ulaganathan, T., Boniecki, M.T., Foran, E., Buravenkov, V., Mizrachi, N., Banin, E., Helbert, W., Cygler, M. (2017). New Ulvan-Degrading Polysaccharide Lyase Family: Structure and Catalytic

Mechanism Suggests Convergent Evolution of Active Site Architecture. ACS Chem. Biol. *12*, 1269–1280.

Ulaganathan, T., Helbert, W., Kopel, M., Banin, E., Cygler, M. (2018). Structure–function analyses of a PL24 family ulvan lyase reveal key features and suggest its catalytic mechanism. J. Biol. Chem. *293*, 4026-4036.

Ulaganathan, T., Banin, E., Helbert, W., Cygler, M. (2018). Structural and functional characterization of PL28 family ulvan lyase NLR48 from Nonlabens ulvanivorans. J. Biol. Chem. jbc-RA118.

Vuong, T. V., Wilson, D.B. (2010). Glycoside hydrolases: Catalytic base/nucleophile diversity. Biotechnol. Bioeng. *107*, 195–205.

Walden, H. (2010). Selenium incorporation using recombinant techniques. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 352–357.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E.E., Owens, R.J., Stuart, D.I.I., Grimes, J.M. (2006). Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. Structure *14*, 1617–1622.

Wargacki, A.J., Leonard, E., Win, M.N., Regitsky, D.D., Santos, C.N.S., Kim, P.B., Cooper, S.R., Raisner, R.M., Herman, A., Sivitz, A.B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D., Yoshikuni, Y. (2012). An engineered microbial platform for direct biofuel production from brown macroalgae. Science. *335*, 308–313.

Wingfield, P.T. (2015). Overview of the purification of recombinant proteins. Curr. Protoc. Protein Sci. 2015, *80*, 6.1.1-6.1.35.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. *280*, 5705–5736.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. *275*, 1–21.

Xu, F., Dong, F., Wang, P., Cao, H.-Y., Li, C.-Y., Li, P.-Y., Pang, X.-H., Zhang, Y.-Z., Chen, X.-L. (2017). Novel Molecular Insights into the Catalytic Mechanism of Marine Bacterial Alginate Lyase AlyGC from Polysaccharide Lyase Family 6. J. Biol. Chem. *292*, 4457–4468.

Xu, F., Wang, P., Zhang, Y., Chen, X. (2018). Diversity of Three-Dimensional Structures and Catalytic Mechanisms of Alginate Lyases. *84*, 1–12.

Yadav, S., Yadav, P.K., Yadav, D., Yadav, K.D.S. (2009). Pectin lyase: A review. Process Biochem. *44*, 1-10.

Yamasaki, M., Moriwaki, S., Miyake, O., Hashimoto, W., Murata, K., Mikami, B. (2004). Structure and function of a hypothetical Pseudomonas aeruginosa protein PA1167 classified into family PL-7: A novel alginate lyase with a β-sandwich fold. J. Biol. Chem. *279*, 31863–31872.

Yang, L.W., Bahar, I. (2005). Coupling between catalytic site and collective dynamics: A

requirement for mechanochemical activity of enzymes. Structure *13*, 893–904.

Yoder, M.D., Keen, N.T., Jurnak, F. (1993). New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. Sci. (New York, NY) *260*, 1503–1507.

Yoder, M.D., Lietzke, S.E., Jurnak, F. (1993). Unusual structural features in the parallel β-helix in pectate lyases. Structure *1*, 241–251.

Yoon, H.J., Hashimoto, W., Miyake, O., Murata, K., Mikami, B. (2001). Crystal structure of alginate lyase A1-III complexed with trisaccharide product at 2.0 Å resolution. J. Mol. Biol. *307*, 9–16.

Yoon, H.J., Mikami, B., Hashimoto, W., Murata, K. (1999). Crystal structure of alginate lyase A1-III from Sphingomonas species A1 at 1.78 Å resolution. J. Mol. Biol. *290*, 505–514.

Zhang, Z.H., Bharatham, K., Sherman, W.A., Mihalek, I. (2010). deconSTRUCT: General purpose protein database search on the substructure level. Nucleic Acids Res. *38*, 590-594.

Appendix

Copyright Waiver

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jul 09, 2018

This Agreement between Mrs. ThirumalaiSelvi Ulaganathan ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4384970817126
License date	Jul 09, 2018
Licensed content publisher	Oxford University Press
Licensed content publication	Glycobiology
Licensed content title	Conformational flexibility of PL12 family heparinases: structure and substrate specificity of heparinase III from *Bacteroides thetaiotaomicron* (BT4657)
Licensed content author	Ulaganathan, ThirumalaiSelvi; Shi, Rong
Licensed content date	Jan 5, 2017
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Structural studies on enzymes involved in uronic acid polysaccharide degradation
Publisher of your work	University of Saskatchewan
Expected publication date	Jul 2018
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Structural studies on enzymes involved in uronic acid polysaccharide degradation
Instructor name	Dr.Miroslaw Cygler
Institution name	University of Saskatchewan
Expected presentation date	Jul 2018
Portions	Text extract, Figures and table
Requestor Location	Mrs. ThirumalaiSelvi Ulaganathan 3D40, Health science building 107 wiggins road Saskatoon, SK S7N5E5 Canada Attn: Mrs. ThirumalaiSelvi Ulaganathan
Publisher Tax ID	GB125506730
Billing Type	Invoice
Billing Address	Mrs. ThirumalaiSelvi Ulaganathan 3D40, Health science building 107 wiggins road Saskatoon, SK S7N5E5

Canada
Attn: Mrs. ThirumalaiSelvi Ulaganathan

Total 0.00 USD

Terms and Conditions

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.
5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employs and agents, from and against any and all claims

arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Other Terms and Conditions:
v1.4

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Editorial policies

JBC jbc.org/site/misc/edpolicy.xhtml

Copyright and license to publish

For authors:

Effective with initial submissions from January 1, 2018, authors must agree that if their manuscript is accepted, they will grant ASBMB an exclusive, irrevocable License to Publish their work; the copyright remains with the authors.

Manuscripts initially submitted prior to January 1, 2018 are subject to JBC's former policy whereby, as a condition of publication, authors transfer copyright to ASBMB upon acceptance.

Authors of manuscripts, submitted at any time, need not contact the journal to request permission to reuse their own material. Authors who granted ASBMB exclusive license to publish and authors who transferred copyright to ASBMB, are allowed to do the following:

a. to use all or part of the work in compilations or other publications of the Authors' own commercial and noncommercial works (includes theses/dissertations), to use figures, photos, and tables created by them and contained in the work, to present the work orally in its entirety, and to make copies of all or part of the work for the Authors' use for lectures, classroom instruction or similar uses. If the author is employed by an academic institution, that institution also may reproduce the article for teaching purposes.
b. to post the accepted manuscript version of the work, the "Paper in Press," on the author's personal web page, their personal or institutional repository, or their funding body's archive or designated noncommercial repository, provided that a link to the article in the *Journal of Biological Chemistry* is included.
c. to post a manuscript version of the work on not-for-profit preprint servers provided that the Authors retain distribution rights to the work, that ASBMB formatted final files are not posted, and that a link to the article in the *Journal of Biological Chemistry* is included.
d. to post the final edited PDFs, created by ASBMB, to their own departmental/university websites, provided that the posting does not happen until 12 months after publication of the work in the *Journal of Biological Chemistry*, and that a link to the article in the *Journal of Biological Chemistry* is included.

Reuse of JBC content must include the following: This research was originally published in the Journal of Biological Chemistry. Author(s). Title. J. Biol. Chem. Year; Vol:pp-pp. © the American Society for Biochemistry and Molecular Biology *or* © the Author(s).